The Dissertation Committee for Philip James Capin certifies that this is the approved version of the following dissertation:

Examining the Role of Treatment Adherence and Instructional Quality in the Context of a Content-Area Reading Instructional Treatment for Eighth Graders

Committee:

Sharon Vaughn, Supervisor

Nathan Clemens

Greg Roberts

Jessica Toste

Examining the Role of Treatment Adherence and Instructional Quality in the Context of a Content-Area Reading Instructional Treatment for Eighth Graders

by

Philip James Capin

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2018

Acknowledgements

Kell, thank you for your support, patience, love, friendship, and most of all belief—your enduring belief in me and in us. It turns out you are the best thing.

Pops, I still hope I am just like you when I grow up. Thank you for your example.

Mom, I feel your support every day. Thank you for being my untiring cheerleader.

Mandy, thank you for helping me to graduate high school. I am not positive, but I am pretty sure that not graduating high school might have been an impediment to this whole graduate school thing.

Scotty, thank you for choosing to be a parent to me and inspiring me to see what you see in me.

Sharon, I was recently reminded how fortunate I am when I read rates of depression and anxiety reported by postgraduate students are six times higher than those reported by people in the general population (Evans et al., 2018). Thank you for always supporting your students to seek work-life balance and happiness in and outside of work. Thank you for encouraging us to start families and put them first. Thank you for always treating your graduate students like smart, hard-working colleagues. Thank you for repeatedly telling us, "Look out for your own career goals. Our research center will be fine, but you need to focus on what is going to help you in your career long-term." Thank you for encouraging us to consider jobs outside of academia if they better meet our goals. Thank you for teaching us how others may define success, but always encouraging us to create our own goals and judge our success by only these standards. Thank you for everything.

Greg, thank you for the many hours of guidance and support you provided over the

last few years. (Please know this acknowledgement means we are square.)

Examining the Role of Treatment Adherence and Instructional Quality in the Context of a Content-Area Reading Instructional Treatment for Eighth Graders

Philip James Capin, Ph.D.
The University of Texas at Austin, 2018

Supervisor: Sharon Vaughn

Relatively few reading intervention studies have investigated treatment fidelity data and the extent to which fidelity is associated with student outcomes. This study extends the extant research by examining treatment adherence and quality and the extent to which these components of treatment fidelity relate to posttest performance for eighth grade students who received a content-area reading treatment targeting improved content knowledge and reading comprehension. Data were analyzed to address the following questions: (1) Do treatment adherence and quality indicate a general factor of fidelity? (2) Do treatment adherence and quality predict greater performance in content acquisition and reading comprehension? (3) To what extent are the effects of treatment adherence and quality on student outcomes moderated by teacher's classroom management performance and student characteristics (pretest reading performance level, limited English proficiency status, special education status, and gender)? Situated within a larger randomized controlled trial examining the effects of a content-area instructional program, this study used data from the subsample of students (n = 790) in classes randomly assigned to treatment. Results showed

a general factor of fidelity did not sufficiently predict treatment adherence and quality. Treatment adherence and quality were significant, positive predictors of content knowledge performance, which indicates that increases in both dimensions of treatment fidelity were associated with improvements in content knowledge posttest performance. However, a significant interaction between treatment adherence and quality was present. Interaction plots revealed instructional quality led to greater levels of posttest performance for students in classrooms with high levels of instructional adherence; however, instructional quality was not strongly associated with posttest performance when instruction was implemented with low adherence. Results also showed instructional quality was a significant, positive predictor of performance on the ASK Reading Comprehension subtest. Moderator analyses revealed teachers' classroom management and students' pretest performance influenced the effects of treatment adherence and quality on student outcomes. These findings help to elucidate the conditions under which the PACT instructional treatment leads to improved outcomes and, more broadly, underscores the importance of examining dimensions of treatment fidelity and the interplay between treatment fidelity dimensions and other teacher- and student-level characteristics when interpreting treatment effects.

Table of Contents

List of Tables

List of Figures

Chapter I: Introduction

Middle and high school content-area teachers are responsible for developing students' content-area knowledge and reading comprehension skills. Over the past decade, expectations for secondary content-area teachers have changed with the adoption of progressive state standards, such as the Common Core State Standards (CCSS; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Progressive state standards call for students to read and make sense of content area texts as well as show adequate content knowledge. These standards require students to be able to (a) understand content-specific vocabulary and text features, (b) identify key ideas and details in text, (c) integrate knowledge ideas across text sources, and (d) and analyze the author's point of view and respond to conflicting viewpoints.

Content-area texts are substantially more challenging than narrative text (Berman & Nir-Sagiv, 2007; De Oliveira, 2010; Harmon, Hedrick, & Fox, 2000) and the development of basic reading skills in the early grades does not guarantee students will be able to read with understanding the complex texts across content areas texts in the secondary grades (e.g., Carnegie Council on Advancing Adolescent Literacy, 2010; Snow, Porche, Tabors, & Harris, 2007). Content-area texts in the secondary grades comprise complex vocabulary, syntax, and text structures (Lee & Spratley, 2010; Shanahan & Barr, 1995). Shanahan and Shanahan (2008) suggest that students receive instruction and practice opportunities to develop the specialized background knowledge and skills needed to understand the complex texts. This suggestion aligns with the progressive state standards calling for integrating reading instruction within content instruction.

Although research findings suggest it is quite difficult to meaningfully impact

reading outcomes for secondary students (e.g., Cheung, Slavin, Lake & Kim, 2016; Kemple, Corrin, Nelson, & Salinger, 2008), there are a few approaches to improving adolescent literacy outcomes grounded in research to guide content-area teachers who seek to integrate reading and content instruction. In an Institute of Education Sciences (IES) practice guide, Kamil and colleagues (2008) make recommendations for improving adolescent literacy outcomes based on the best available evidence. These recommendations advocate for classroom teachers to (a) provide explicit vocabulary and comprehension strategy instruction, (b) offer opportunities for students to engage in extended discussions of text meaning, (c) enhance student motivation and engagement, and (d) deliver intensive interventions for struggling readers. These recommendations align with the findings of meta-analyses for students in the secondary grades examining the effects of reading interventions for secondary struggling readers (Scammacca, Roberts, Vaughn, & Steubing, 2015) and students with disabilities (Berkeley, Scruggs, & Mastropieri, 2010). Specifically, these meta-analyses also highlight the positive effects associated with delivering explicit strategy instruction, incorporating direct vocabulary instruction, and providing students multiple opportunities to practice reading and understand texts.

Despite heightened standards for adolescent content learning and literacy and converging guidance in policy reports about effective practices for improving adolescent literacy (e.g., Biancarosa & Snow, 2006; Kamil et al., 2008; Torgesen et al., 2007), national and international achievement data show significant numbers of adolescents in U.S. schools do not adequately understand the texts they are asked to read (U.S. Department of Education, 2015; Organization for Economic Cooperation and Development [OECD], 2016). The most recent data from the National Assessment of Educational Progress

(NAEP) show that only 34% of eighth grade students are proficient in reading (U.S. Department of Education, 2015). International data from the Program for International Student Assessment (PISA) show that 15-year-olds in the U.S. perform more poorly than their peers in 22 countries (OECD, 2016). Reading comprehension deficits are of particular concern given inadequate reading skills are a primary cause of academic failure at the secondary level (e.g., Biancarosa & Snow, 2004) and an impediment to pursuing postsecondary education and finding success in the workforce (e.g., Kamil et al., 2008). This is especially problematic for secondary content area teachers as students are expected to learn content by reading texts (Toste, Fuchs, & Fuchs, 2013). Hence, it is not unexpected that national data also show that U.S. students underperform in content area subject knowledge, such as history and science. For instance, only 18% of eighth graders are proficient in history and social studies (U.S. Department of Education, 2015). Importantly, neither the percentage of students scoring proficient in 2015 in the area of reading, nor the percentage of students meeting proficiency in U.S. history, increased a statistically significant amount since the previous testing period (U.S. Department of Education, 2015).

Observational data collected with secondary teachers may help to shed light on why adolescents are struggling in the area of reading comprehension and content learning. Observational studies show teachers spend little time integrating comprehension instruction into their content instruction (Pressley, 2004; Swanson et al., 2016). For instance, Swanson et al. (2016) found comprehension instruction was present in only 20% of social studies classrooms. Moreover, Swanson, Wexler, and Vaughn (2009) found that students spent less than 10% of instructional time engaging with text.

Why do teachers infrequently integrate reading comprehension instruction within

content-area lessons despite the need to improve reading comprehension among secondary students? There are a few possible explanations. Although recommendations in policy reports call for all secondary content teachers to possess a foundational understanding of reading instruction to address the needs of adolescent readers (Carnegie Council on Advancing Adolescent Literacy, 2010; International Reading Association, 2012; National Institutes of Literacy, 2007; RAND Reading Study Group, 2002), preparation programs for secondary content-area teachers spend little time preparing teachers to provide reading instruction (Ness, 2009). Additionally, many secondary schools do not have reading specialists who can support content-area teachers in delivering reading instruction (Anders, 2002; Snow et al., 2005). Thus, as supported by interview data, secondary teachers generally do not feel qualified to teach their students how to comprehend texts (Anders, Hoffman, & Duffy, 2000; Ness, 2009).

Recognizing the considerable challenges facing teachers, the IES made a substantial investment in the identification of effective practices for improving reading comprehension through the Research for Understanding Research Initiative (Douglas & Albro, 2014). This initiative involved six research teams, over 130 researchers, and a commitment from the IES of over 110 million dollars. These studies improved researchers understanding of the processes involved in reading comprehension and ways to remediate difficulties in this area. One promising approach for improving secondary students' reading comprehension and content knowledge developed and tested through the initiative is Promoting Adolescents' Comprehension of Text (PACT; Vaughn et al., 2013).

Anchored in text-processing theories of reading comprehension (Kintsch, 1974; Pressley, Symons, Snyder, & Cariglia-Bull, 1989), PACT is a content-based approach that

uses discussions about text as the means for improving reading comprehension. Working in partnership, researchers and expert teachers developed and revised the program using design experiments to ensure the program was feasible for teachers to implement and engaging for students. Two large-scale experimental studies found that PACT was consistently associated with positive outcomes on a psychometrically-sound measure of social studies content knowledge (ES = .17 to .32; Vaughn et al., 2013; Vaughn et al., 2015) and one study found the PACT intervention also had a significant effect on students' reading comprehension (ES = .20) as assessed by a broad measure of reading comprehension and a measure of reading comprehension with social studies texts (Vaughn et al., 2013). Moreover, a quasi-experimental study showed the effects of PACT were equivalent or greater for students with disabilities on a measure of content knowledge (ES = .26) and reading comprehension using social studies texts (ES = .34; Swanson, Wanzek, Vaughn, Roberts, & Fall, 2015).

A study examining the effects of PACT using a within-teacher, class-randomized design analyzed the extent to which treatment fidelity mediated the effect of treatment on students' content knowledge and reading comprehension (Vaughn et al., 2015). The authors used treatment adherence data (extent to which a treatment was implemented as intended) to understand the relation between treatment assignment and student outcomes. Specifically, the authors documented the presence of treatment components teachers delivered to treatment (i.e., classes provided PACT instruction) and comparison classes to index treatment differentiation. Their data suggest treatment adherence was relatively high in the treatment classes and there was little crossover to comparison classes. Moreover, the authors found that treatment adherence mediated the effect of treatment on content

knowledge, strengthening the authors' claim that the independent variable was responsible for changes in content knowledge. This dissertation will expand upon this previous work examining treatment fidelity in the context of the PACT instructional practices in a few ways. For one, this study will investigate the relation between instructional adherence as well as instructional quality and content knowledge and reading comprehension performance. Second, this study will focus on explaining variation among the students in classes randomized to receive the PACT program to better understand the circumstances under which PACT leads to improve outcomes. Third, this paper will explore potential moderators of the effects of treatment fidelity on student outcomes, such as teacher's classroom management and student-level characteristics (e.g., special education status). In so doing, this study aims to better understand the mechanisms through which the PACT program leads to improvements in student outcomes and investigate whether there is empirical support for measuring multiple dimensions of treatment fidelity.

Treatment Fidelity

Researchers have historically regarded treatment fidelity (sometimes referred to as fidelity of implementation) as an important methodological facet (e.g., Gresham, 1989; Moncher & Prinz, 1991). Thorough consideration and reporting of treatment fidelity is critical to the internal and external validity of intervention studies (Roberts, Vaughn, Beretvas, & Wong, 2017). Studies that report treatment fidelity provide greater confidence that the results are attributed to the manipulation of the independent variable (Hohmann & Shear, 2002). Conversely, ignoring treatment fidelity in intervention studies can lead to inaccurate study conclusions. For example, consider a study that reports nonsignificant group differences and omits treatment fidelity information. It is not possible to discern in

this case whether the nonsignificant differences were a function of an ineffective treatment or an effective program poorly implemented (Moncher & Prinz, 1991). Thus, researchers who ignore treatment fidelity may abandon a treatment that may be efficacious under other circumstances. Likewise, researchers who find significant differences favoring a treatment condition and do not consider treatment fidelity have overlooked an opportunity to strengthen their causal claim that change in the dependent variable was due to the independent variable rather than an unseen factor (Cook & Campbell, 1979). Treatment fidelity data also facilitate conclusions regarding the external validity of intervention studies. Detailed descriptions of treatment fidelity may help future researchers consider how findings may generalize to other populations and settings in which treatment fidelity may differ from the original experiment. In addition to providing information that affects the validity of an intervention study, research shows the effects of treatment are, on average, higher when treatment fidelity is higher (Wilson & Lipsey, 2001).

Although treatment fidelity is crucial to the validity of intervention studies, researchers across disciplines and areas of research underreport treatment fidelity in intervention studies (Gearing et al., 2011). For example, reviews of psychotherapy (Perepletchikova, Treat, & Kazdin, 2007) and anger management (Gansle, 2005) intervention studies found treatment fidelity was reported in only 3.5% and 10% of studies, respectively. In the area of school-based interventions, reviews of fidelity reporting show school-based intervention studies inconsistently report treatment fidelity (Capin, Walker, Vaughn, & Wanzek, 2017; Gresham, Gansle, & Noell, 1993; McIntyre, Gresham, DiGennaro, & Reed, 2007; O'Donnell, 2008, Swanson, Wanzek, Haring, Ciullo, & McCulley, 2011; Wheeler, Baggett, Fox, & Blevins, 2006). For instance, Capin et al.

(2017) found less than half of kindergarten through Grade 3 (K-3) reading intervention studies conducted from 1995 through 2015 reported treatment fidelity. Additionally, reviews show intervention researchers rarely use treatment fidelity data in their analyses of treatment effects (Capin et al., 2017; National Research Council; O'Donnell, 2008). For instance, O'Donnell (2008) located only five studies in a review of K-12 curriculum interventions that measured the relation between treatment fidelity and student outcomes.

Researchers have noticed an uptick in the frequency of fidelity reporting over time (Capin et al., 2017; Swanson et al., 2011). This trend is likely to continue given the requirements for treatment fidelity reporting by funding agencies (National Institutes of Health, 2011; U.S. Department of Education, 2011) and position statements from professional organizations (e.g., National Association of School Psychologists, 2005) highlighting the importance of analyzing and reporting fidelity information. Despite an increasing rate of fidelity reporting over time, previous syntheses suggest intervention studies could improve their methodology by measuring multiple dimensions of treatment fidelity (e.g., adherence to instructional components, instructional quality, and dosage; Capin et al., 2017; O'Donnell, 2008) and using these data in analyses of treatment effects (Capin et al., 2017; Swanson et al., 2011).

Although it is uncommon for studies to examine the relation between treatment fidelity and student outcomes, a few reading intervention studies in recent years (e.g., Benner, Nelson, Stage, & Ralston, 2011; Fogarty et al., 2014) have explored this relation in an effort to unpack the "black box" to better understand why treatments are or are not effective. Several studies conducted in the last decade have explored multiple dimensions (e.g., adherence and quality) of treatment fidelity in their analyses and used these results to

interpret study findings. For instance, Benner et al. (2011) examined treatment adherence and quality of instruction to determine whether they accounted for variance in the effects of the Corrective Reading Decoding (Engelmann, Hanner, & Johnson, 2002) program for middle school students with reading difficulties. Their findings showed these dimensions did account for significant variance in reading outcomes, adding strength to the claim that improved outcomes were a function of the reading treatment. Fogarty et al. (2014) investigated the relation between fidelity and student outcomes in light of non-significant group differences on reading measures. Results showed treatment fidelity measured as a composite variable comprised of adherence, quality, dosage, program differentiation, and student responsiveness was significantly related to outcomes, indicating high levels of implementation led to greater gains in student outcomes.

Summary

There is a need to identify effective and feasible ways for secondary teachers to improve reading comprehension and content knowledge among adolescents during content-area instruction. PACT is a theoretically-grounded, evidence-based approach to improving content learning and reading comprehension. A previous examination of PACT shows treatment adherence mediated the effect of treatment on students' acquisition of content and reading comprehension (Vaughn et al., 2015). Conceptual papers (e.g., Dane & Schneider, 1998) and emerging empirical research (e.g., Fogarty et al., 2014) suggest future research consider multiple dimensions of treatment fidelity and the role they play in explaining student outcomes. Understanding treatment fidelity may be particularly important in the area of secondary content-area reading instruction given the changing expectations for content-area teachers and the results of previous descriptive research

showing secondary content-area teachers do not feel qualified to teach reading (Anders, Hoffman, & Duffy, 2000; Ness, 2009).

Research Questions

This study added to the research base on treatment fidelity by conducting secondary analyses examining how treatment adherence and instructional quality predict student outcomes using extant data from a large-scale randomized controlled trial of PACT (Vaughn., 2015). The following research questions were addressed:

RQ1. Do treatment adherence and quality indicate a general factor of fidelity?

RQ2: Do treatment adherence and quality predict greater performance in content acquisition and reading comprehension for students who received a content-area reading treatment?

RQ3: To what extent are the effects of treatment adherence and quality on student outcomes moderated by teacher's classroom management performance and student characteristics (i.e., pretest reading performance level, limited English proficiency status, special education status, gender)?

Chapter 2: Review of Previous Research

The goal of this chapter is to define key terms and concepts related to treatment fidelity, as well as review the extant research that informed this secondary analysis of treatment fidelity within a reading intervention study. Specifically, this chapter (a) defines treatment fidelity, (b) describes the current recommendations for measuring and analyzing treatment fidelity data, (c) reviews previous syntheses of treatment fidelity reporting in intervention research, and (d) highlights the most recent intervention research using relatively innovative and comprehensive approaches to examine the effects of treatment fidelity.

Defining Treatment Fidelity

The concept of treatment fidelity has grown more sophisticated with time. Treatment fidelity was initially referred to as treatment integrity and described as the degree to which a treatment is implemented as intended (Yeaton & Sechrest, 1981). Although treatment fidelity is still generally defined as the extent to which a treatment is implemented as intended, the concept has expanded to include several dimensions (or subtypes) of treatment fidelity. Kazdin (1986) and Moncher and Prinz (1991) expanded the concept of treatment fidelity to include treatment differentiation, which involves whether the study conditions differ from one another in the intended manner. Building upon this novel work, Dane and Schneider (1998) identified and defined five dimensions of treatment fidelity (paraphrased below).

1. Adherence (also referred to as treatment adherence): extent to which critical components are implemented as prescribed;
2. Quality: measure of instructional quality separate from adherence;

3. Exposure (or dosage): amount of instruction provided;
4. Participant responsiveness (sometimes referred to as treatment receipt): extent to which participants responded to the intended treatment as indicated by participation or engagement;
5. Program differentiation: extent to which the treatment varies from the comparison condition.

An increasing number of researchers (e.g., Dusenbury, Brannigan, Falco, & Hansen, 2003; Fogarty et al., 2014) have begun to adopt the multidimensional framework of treatment fidelity proposed by Dane and Schneider (1998); however, there is not universal agreement about the dimensions of treatment fidelity. For instance, many intervention studies continue to conceptualize treatment fidelity as solely adherence to treatment components (Gearing et al., 2011; Dusenbury et al., 2003; O'Donnell, 2008). To complicate matters further, a variety of interchangeable terms have been used to describe treatment fidelity, including fidelity, fidelity of implementation, treatment integrity, and intervention fidelity (Breitenstein et al., 2010; Roberts et al., 2017).

It is also important to note the term fidelity often carries a different meaning for practitioners and researchers. For school practitioners, fidelity refers to whether a curriculum or program is executed in the same way it was designed to be used. Teachers considering the fidelity of implementation may ask questions such as, were the important program procedures followed? Were the required number of lessons provided? For researchers working in an experimental context, the focus is on these elements of fidelity as well as whether the treatment condition was implemented as intended relative to a comparison condition. Understanding treatment fidelity in an experimental context enables

researchers to have greater confidence that the experimental differences are a function of the treatment. Thus, researchers may also ask, to what degree did the treatment condition vary from the comparison condition?

For the purposes of this study, I use the term treatment fidelity and adopt the multidimensional framework put forth by Dane and Schneider (1998) for explaining treatment fidelity. Further, I consider treatment fidelity in an experimental context, which requires not only understanding variation that exists within treatment implementation but also collecting information about the counterfactual to index treatment differentiation.

Recommendations for Treatment Fidelity in Intervention Studies

As the concept of treatment fidelity has evolved over time, the recommendations for defining, enhancing, measuring, and analyzing treatment fidelity have also grown more complex. Moncher and Prinz (1991) proposed the first set of specific guidelines for the enhancement of treatment fidelity. These guidelines encouraged researchers to (a) operationally define the treatment, (b) adequately train implementers for treatment delivery using treatment manuals, (c) provide ongoing supervision to treatment implementers, (d) measure adherence to treatment via outside observations, and (e) utilize fidelity data to interpret research findings. Many of these early recommendations for treatment fidelity remain central to more recent guidelines put forth by health scientists (e.g., Bellg et al., 2004; Borelli et al., 2005; Miller & Rollnick, 2014) and educational researchers (e.g., O'Donnell, 2008; Swanson et al., 2011).

Building on the original work of Moncher and Prinz (1991), health scientists have advanced the definition, monitoring, and measurement of treatment fidelity over the past 20 years. Key to these advancements was the formation of the National Institutes of Health

(NIH) Treatment Fidelity Workgroup in 1999, as part of a consortium of NIH-funded projects focused on improving health-related outcomes. This workgroup put forth several recommendations for researchers related to best practices in treatment fidelity: (1) train and supervise treatment implementers using specific strategies, and assess implementers before treatment delivery to ensure implementers acquire critical skills; (2) measure not only treatment adherence and dosage, but also variation in treatment fidelity among implementers, treatment differentiation, and treatment receipt; (3) collect data on treatment and comparison sessions using audio tapes or observations, and conduct exit interviews with comparison group implementers to examine treatment differentiation; and (4) monitor and describe treatment receipt to ensure participants understand and can make use of health-related treatments (Bellg et al., 2004).

Special education researchers have also acknowledged the important role of treatment fidelity and offer recommendations for treatment fidelity reporting in educational research. Recommendations for reporting treatment fidelity were published in *Exceptional Children* for experimental and quasi-experimental group design studies (Gersten et al., 2005) and single case studies (Horner et al., 2005) in special education. Describing the treatment fidelity quality indicators for group studies, Gersten and colleagues (2005) asserted that assessing and reporting treatment adherence is an "essential" element of high-quality studies (p. 156). These guidelines specified that authors conduct regular instructional observations and use a checklist of the most important features of the intervention to measure adherence. Similarly, Horner and colleagues (2005, p. 168) called for the "continuous and direct measurement" of treatment fidelity within single case studies. Although not posited as a requirement, Gersten and colleagues (2005, p.156)

considered assessing and describing quality of implementation (i.e., how well the treatment was implemented) as "desirable" for evidence-based practices high-quality group studies in special education. Adherence and quality of intervention are particularly worthy of measurement in educational research given implementers of educational treatments are often required to enact instructional practices without a scripted protocol and make countless instructional decisions in short periods of time (Mowbray, Holter, Teague, & Bybee, 2003).

Other educational researchers have suggested that authors investigate the technical properties of fidelity measures and use these measures to examine the relation between treatment fidelity and student outcomes. O'Donnell (2008) recommended authors describe how fidelity measures were constructed and both assess and report the reliability and validity of the treatment fidelity data collected (Mowbray, Holter, Teague, & Bybee, 2003; O'Donnell, 2008). Mowbray and colleagues (2003) identified several approaches to assessing reliability and validity of fidelity measures, including calculating inter-rater agreement across fidelity raters (reliability) and examining the agreement between two different sources of fidelity information (validity). Additionally, experts have suggested researchers examine the relation between treatment fidelity and student outcomes based on previous research showing treatment fidelity data were associated student outcomes (e.g., National Research Council, 2004; O'Donnell, 2008).

Although researchers across disciplines suggest defining and measuring treatment fidelity slightly differently, continuing efforts to advance treatment fidelity research provide educational researchers with an array of methodological considerations to help unpack issues related to treatment implementation. Taken together, current guidelines

suggest researchers consider (a) processes for training, supporting, and assessing treatment implementers; (b) procedures for collecting fidelity data, and (c) techniques for analyzing the relation between treatment fidelity and outcomes.

Treatment Fidelity Reporting in Educational Intervention Research

A few trends are present in the extant reviews of treatment fidelity in educational intervention research. First, school-based interventions inconsistently report treatment fidelity information in intervention studies (Capin et al., 2017; Gresham, Gansle, & Noell, 1993; McIntyre et al., 2007; O'Donnell, 2008, Swanson et al., 2011; Wheeler et al., 2006), yet the frequency of reporting has increased over time (Capin et al., 2017; Swanson et al., 2011). For instance, in the field of learning disabilities, a review of treatment fidelity reporting on academic intervention studies published in select special education journals from 1995 to 1999 found only 18% of studies reported treatment fidelity (Gresham et al., 2000). Swanson and colleagues (2011) reviewed the five highest impact general and special education journals that published intervention studies from 2005 to 2009, finding that 47% of studies reported treatment fidelity scores (with similar levels of reporting in both general and special education journals). This demonstrated a notable increase in treatment fidelity reporting relative to the previous learning disabilities synthesis (Gresham et al., 2000). In the most recent investigation of treatment fidelity reporting, Capin and colleagues (2017) examined treatment fidelity reporting in a corpus of 175 K-3 reading intervention studies published from 1995 to 2015. Results showed less than half (47%) of reading intervention studies synthesized reported treatment fidelity; however, there was a significant association between study year and proportion of fidelity reporting, indicating treatment fidelity reporting was improving over time.

Another trend apparent in past reviews of treatment fidelity is that they did not often collect data for all dimensions of the construct (Capin et al., 2017; O'Donnell, 2008). For example, most school-based treatment fidelity syntheses limited the construct of fidelity to adherence, omitting other fidelity dimensions such as dosage, quality of implementation, treatment receipt, and treatment differentiation (McIntyre et al., 2007; Wheeler, Baggett, Fox, & Blevins, 2006). Swanson and colleagues (2011) examined treatment adherence, quality of implementation, and dosage; however, did not address treatment receipt and treatment differentiation in their synthesis. Capin et al. (2017) found that 72% of the studies that reported treatment fidelity only reported treatment adherence. Moreover, only two studies in the corpus of 175 reading intervention studies evaluated treatment differentiation, an essential element to assessing indexing treatment fidelity differences in an experimental environment (Roberts et al., 2017). This is particularly problematic in educational research as studies have shown there is often considerable, unintended overlap between the instruction received by treatment and comparison students (e.g., Lemons, Fuchs, Gilbert, & Fuchs, 2014; Vaughn, Solis, Miciak, Taylor, & Fletcher, 2016).

A final common finding from examinations of the previously conducted intervention studies is that authors infrequently use treatment fidelity data to analyze treatment effects (Capin et al., 2017; O'Donnell, 2008; Swanson, 2011). In education, O'Donnell (2008) located only five studies that statistically measured the relation between treatment fidelity and outcomes among K–12 studies. All five of the studies identified by O'Donnell (2008) reported that higher treatment fidelity was associated with statistically significantly improved student outcomes. Swanson and colleagues (2011) also investigated

how authors incorporated fidelity data into the analysis of intervention effects. Of the 50 studies that reported treatment fidelity data, only two studies used fidelity data to interpret conclusions. Correspondingly, Capin et al. (2017) found only four of the 83 studies reporting treatment fidelity (5%) used this data in their analyses of treatment effects.

Advancements in the Use of Treatment Fidelity to Understand Treatment Effects

Substantial conceptual work has recommended the examination of treatment fidelity in advanced ways. Adopting recommended practices has the potential to strengthen the validity of educational intervention studies and help researchers and practitioners understand the active ingredients within a treatment and the conditions under which these practices enhance student outcomes (e.g., Dane & Schneider, 1998; Kazdin, 1986). Some researchers have begun to explore the dimensionality of treatment fidelity and scrutinize the role treatment fidelity plays in student outcomes in intervention research. Yet, until recently, empirical support for the long-standing conceptual assertions regarding treatment fidelity were largely absent from the existing research (Vaughn et al., 2015). In recent years, there appears to be a shift from only descriptively reporting treatment fidelity on a single dimension of fidelity to more innovative and comprehensive approaches of exploring multiple dimensions of treatment fidelity for units assigned to treatment and comparison conditions using statistical methods. Indeed, at least ten studies published over the last decade have measured multiple dimensions of treatment fidelity and analyzed the relation between treatment fidelity and student academic outcomes using statistical techniques (Benner, Nelson, Stage, & Ralston, 2011; Buckley, Moore, Boardman, Arya, & Maul, 2017; Boardman, Klingner, Buckley, Annamma, & Lasser, 2015; Domitrovich et al., 2010; Fogarty et al., 2014; Guo et al., 2016; Hamre et al., 2010; Justice, Kaderavek, Fan, Sofka,

& Hunt, 2009; Mendive et al., 2016; Vaughn et al., 2015). This research, although nascent, presents preliminary empirical support to the conceptual work that has served as the foundation for study into treatment fidelity.

Several of the studies investigating the relations between treatment fidelity and student outcomes include treatment fidelity as a central consideration in the interpretation of study results. For example, in a study examining the effects of a multicomponent reading comprehension intervention for middle school students, Fogarty et al. (2014) found that, on average, there were no significant between-group differences on targeted outcomes. Yet, the authors utilized data on multiple dimensions of treatment fidelity to show the intervention led to significant, although modest, gains when high levels of treatment adherence, quality, dosage, differentiation, and receipt were present. Mendive, Weiland, Yoshikawa, and Snow (2016) examined the effects of a professional development program implemented with pre-K and kindergarten teachers on reading outcomes and, similar to Fogarty et al., results showed no significant between-group differences on intended outcomes. To better understand the null results, Mendive et al. delved into their treatment fidelity data to understand whether the professional development program was ineffective in changing teacher behavior or changes in teacher behavior did not lead to improved outcomes. Results showed there was a significant association between dosage and pre-literacy outcomes; however, the professional development program led to only modest differences in the amount of time teachers spent on important instructional elements (dosage). Mendive et al. interpreted these findings as evidence that the professional development program failed to impact teaching practice to a large enough degree to produce substantial effects on student outcomes. These studies and others (e.g., Vaughn et

al., 2015) suggest analyses of treatment fidelity can influence decisions about whether a program is effective or not and provide direction to researchers developing instructional programs to be tested in less controlled settings.

The growing corpus of studies that report on multiple dimensions of fidelity and relating these data to outcomes can also help to unpack the important elements of treatment programs. One illustrative example comes from an early childhood education language and pre-literacy study conducted by Hamre et al. (2010). Hamre et al. examined three indicators of treatment fidelity—dosage, adherence, and quality of delivery—and their impact on student outcomes. Results noted a strong association between instructional quality and student outcomes; however, adherence was not related to student outcomes and dosage was not strongly related to outcomes. Interestingly, results showed instructional quality was more strongly related to student outcomes for both students who showed the lowest skills at the beginning of the treatment and students who did not speak English at home. These findings and other research examining student by treatment interactions (e.g., Connor, Morrison, & Petrella, 2004) suggests there may be student by treatment fidelity interactions worth exploring in future research.

Despite recent advancements in treatment fidelity, there are at least two ways in which current research on treatment fidelity should be extended. For one, more systematic measurement studies of treatment fidelity are needed to understand the best ways to conceptualize treatment fidelity. Fogarty et al. (2014) conducted a confirmatory factor analysis (CFA) representing fidelity as a single factor that predicts variance in five dimensions of treatment fidelity (i.e., adherence, quality, dosage, responsiveness, and differentiation). This study provides some empirical support for analyzing treatment

fidelity as latent factor with multiple dimensions (Fogarty et al., 2014). Alternatively, Guo et al. (2016) examined the dimensionality of treatment fidelity for an early childhood intervention using CFA to find a three-factor structure that included adherence/dosage, participant responsiveness, and program differentiation yielded the best model fit relative to a single fidelity factor. Thus, it remains an unresolved question as to how treatment fidelity may be best modeled. It is possible the answer may vary by study based on the treatment under investigation and how authors conceptualize and assess treatment fidelity.

A second area for future intervention research involves examining other factors that may influence the relation between treatment fidelity and student outcomes. As previously discussed, Hamre et al. (2010) found student characteristics, namely student's initial status in language and preliteracy skills and home language, influenced the association between treatment fidelity and student outcomes. Specifically, the association between instructional quality and student outcomes was greater for students who did not speak English at home and showed lower initial performance on outcomes of interest. Though not fully addressed in the extant literature, there may be other instructional factors (e.g., teacher's classroom management skills) and student characteristics (e.g., attention to task) that may influence the relation between treatment fidelity.

Summary

Treatment fidelity is a multidimensional construct that describes the extent to which a treatment is implemented as intended. Recognizing that treatment fidelity data are vital to experimental validity (e.g., Shadish et al., 2002) and studies with higher levels of treatment fidelity are associated with greater treatment effects (Wilson & Lipsey, 2001), researchers from related fields have developed guidelines for conceptualizing, measuring, and

analyzing treatment fidelity data in intervention studies. These guidelines call for intervention researchers to (a) consider multiple dimensions of treatment fidelity; (b) take steps to enhance treatment fidelity through implementer training, support, and assessments of implementer knowledge and skills before the onset of treatment; (c) have independent researchers collect treatment fidelity using reliable and valid measures; and (d) analyze the relation between treatment fidelity and outcomes. Notwithstanding converging recommendations from multiple sources, reviews of treatment fidelity in educational intervention research studies show treatment fidelity is frequently omitted from published intervention articles. Moreover, those studies reporting treatment fidelity typically operationalized it as a unidimensional construct and did not collect information about the comparison condition. This is particularly problematic in educational research as studies have shown that there is often considerable, unintended overlap between the instruction received by treatment and comparison students (e.g., Lemons, Fuchs, Gilbert, & Fuchs, 2014; Vaughn, Solis, Miciak, Taylor, & Fletcher, 2016). Additionally, past reviews show it is rare for educational interventions to use treatment fidelity to statistically analyze the relations between assignment to treatment, treatment fidelity, and student outcomes. Finally, the review of relevant literature highlighted some of the innovative methods researchers have utilized to statistically examine the relations between assignment to treatment, treatment fidelity, and student outcomes. The review showed recent studies are beginning to empirically substantiate the assertions put forth in conceptual work on treatment fidelity and highlighted the need for additional measurement studies of treatment fidelity and examination of factors that may influence the relation between treatment fidelity and student outcomes.

Chapter 3: Methodology

Overview

The purpose of this paper was to add to the extant research on treatment fidelity and its relation to student outcomes by conducting secondary analyses examining how treatment adherence and quality predict student outcomes using data from a randomized controlled trial of PACT (Vaughn., 2015). Moreover, these secondary analyses may contribute to understanding the conditions under which PACT leads to improvements in content acquisition and reading comprehension. The following research questions were addressed:

RQ1: Do treatment adherence and quality indicate a general factor of fidelity?

RQ2: Do treatment adherence and quality predict greater performance in content acquisition and reading comprehension for students who received a content-area reading treatment?

RQ3: To what extent are the effects of treatment adherence and quality on student outcomes moderated by teacher's classroom management performance and student characteristics (i.e., pretest reading performance level, limited English proficiency status, special education status, gender)?

Data from this study is derived from a multi-site randomized controlled trial (RCT) that examined the effects of PACT on content knowledge and reading comprehension outcomes for eighth grade students in general education history classrooms. The PACT RCT (Vaughn et al., 2015) employed a within-teacher randomized design in which an independent researcher randomly assigned eighth grade social studies teachers' classes to use treatment instructional practices (PACT) or typical practice instruction. Teachers taught

identical social studies content in both treatment and comparison classrooms over the same period of time. This within-teacher design aimed to control for teacher effects and isolate the effects of instructional delivery (i.e., PACT or typical practice). In total, 7 schools, 19 teachers, 85 classes, and over 1,487 students participated in the RCTs. The current study examined the relation between treatment fidelity and student outcomes for units randomized to treatment. In total, 47 classes and a total of 790 students were randomized to treatment.

This chapter includes a review of the methods employed in the previous PACT RCT (Vaughn et al., 2015), as well as the specific methods proposed for the present study.

Methods

Participants

Nineteen U.S. history teachers from seven economically and ethnically diverse middle schools across five large school districts in the Southwest and Southeast of the U.S. participated in the research studies. Each teacher taught at least two classes and some teachers taught as many as six classes of U.S. History. Of the 19 teachers who participated in the studies, 53% were female, 90% were White, and 10% were Hispanic. All teachers possessed a bachelor's degree and about one-third of the teachers held a master's degree.

The student sample for the present paper included all students from participating eighth grade classrooms who consented to participation and were randomized to treatment. In total, 790 students from 47 U.S. History classes received PACT instruction. The ethnic composition of the treatment sample (51% White, 24% Hispanic, 19% African American, and 6% other or multiple races) resembles the demographics of K-12 students in U.S. public schools. Of the sample, 39% of students received free or reduced lunch, 5% were

designated as limited English proficient, and 8% were identified with a disability. Treatment student demographic data are presented in Table 3.1.

Table 3.1

Student Demographic Data

	N	%
Gender		
Female	387	50 %
Male	388	50 %
Special Education		
No	708	91%
Yes	64	8%
Missing	3	1%
Free/Reduced Lunch		
No	473	61%
Yes	302	39%
Limited English Proficient		
No	733	95%
Yes	42	5%
Race		
Latino/Hispanic	189	24%
Caucasian/White	396	51%
African American/Black	145	19%
Other or Multiple Races	45	6%

Implementation Procedures

Participating teachers completed 8-10 hours of training on the PACT program and the procedures for participating in the research study. During these trainings, the research team provided an expansive description of the research design, procedures, and methods to safeguard against transfer of instructional practices from treatment to comparison conditions. As teachers implemented instruction, the research team provided in-class support visits and met with teachers during after-school planning sessions to support appropriate implementation. In-class support took the forms of modeling, coteaching, and observations with feedback. During after school planning sessions, researchers provided further instructional support by leading brief presentations about important instructional elements and methods for maintaining separation between instructional conditions. Additionally, researchers answered teacher questions and provided teachers an opportunity to discuss challenges and solutions with the other participating teachers at their school site.

Instruction spanned from six to eight weeks and consisted of 30 instructional lessons. Each lesson lasted approximately 50 minutes. As teachers implemented PACT with their classes assigned to receive PACT instruction, teachers simultaneously provided an equal number of lessons about the same historical content using their typical approach to social studies instruction to classes that were randomized to the comparison condition.

Description of PACT

PACT is a content-area reading approach developed to support the acquisition of content knowledge and improvement of reading comprehension among students in the secondary grades. With funding from the Reading for Understanding initiative from the IES, researchers and expert teachers collaborated to develop PACT using design

experiments to ensure the program was feasible and engaging for students. As previously described, PACT is grounded in reading comprehension theories that conceive of reading comprehension as a process that involves constructing coherent representations of text and integrating these representations with prior learning (Kintsch, 1974; van den Broek, Young, Tzeng, & Linderholm, 1998). The PACT approach aligns with text-processing approaches (also referred to as content-approaches; McKeown, Beck, & Blake, 2009) to reading instruction in which teachers direct students to use text to build a representation of the essential ideas and facilitate understanding through discussion. This approach represents a deviation from the strategy approach, which has played a prominent role in reading comprehension research and practice for the past thirty years (Gersten, Fuchs, Williams, & Baker, 2001; Solis et al., 2012).

As noted, PACT teachers implement a set of 30 lessons to all students in U.S. History classes. The lessons are organized in three ten-lesson units that focused on Colonial America, the Road to Revolution, and the Revolutionary War. PACT consists of six key instructional components: (1) comprehension canopy, (2) warm-up, (3) essential words, (4) content acquisition, (5) team-based learning (TBL) comprehension checks, and (6) TBL knowledge application activity.

Comprehension canopy. The comprehension canopy supports students in building background knowledge and motivation about the social studies content. The comprehension canopy step occurs before reading the day's text. At the beginning of each unit, this step involves students watching a short video (3-7 minutes) on the topic. Before showing the video clip, teachers provide students with a purpose for viewing the content. After viewing the video, students engage in a small group or class-wide discussion that

addresses the purpose for viewing, as well as any additional questions raised by teachers or students. Teachers then pose a difficult question related to the content of the entire unit for the students to consider as they read the day's text. At the beginning of each subsequent lesson within the unit, the teachers direct students to the comprehension canopy question (often posted on a classroom wall in view of all students) and provide students with an opportunity to discuss new knowledge that relates to the comprehension question. At the end of each unit, teachers expect students to be able to answer the comprehension canopy question.

Warm up. In each unit, lessons called for teachers to dedicate five to ten minutes during four lessons for warm-up activities to jump-start the day's lesson. During this time, teachers directed students to a warm-up activity in their student handbook, provided them time to complete the activity, and discussed student responses as a group. The warm-up activities integrated previous content and vocabulary learning by relating to the previous day's lessons and incorporated key vocabulary explicitly taught during essential words instruction.

Essential words. Essential words instruction helps students to build vocabulary related to the overarching concepts of each unit. In each unit, four to five high utility words (e.g., protest) central to the contents of the unit are taught using visuals, student-friendly definitions, and opportunities to practice using the words in context. On the first day of each unit, the teacher presents each essential word to the students using a simplified definition, image, related words, and sentences using the word in context before reading the day's text. In each of the subsequent nine lessons, the teacher reviews a couple of the essential words and provides question prompts to students to initiate discussions about the

meaning of words in context while employing turn-and-talk procedures.

Knowledge acquisition. Knowledge acquisition is the essential component of PACT instruction as it establishes a process for using texts as the resource for students to expand knowledge and enhance their ability to comprehend texts. During the knowledge acquisition step, students read text sections (section length will vary based on student ability), identify important ideas and connect these ideas to prior learning, have an opportunity to ask questions to the teacher and their peers and respond to teacher checks for understanding, discuss key ideas and make connections, and take notes. Students can read texts individually, in pairs, in small groups, or as a whole class. Once a text section is read, students are given an opportunity to ask questions, and the teacher will ask a couple of brief comprehension questions to assess whether students have a basic understanding of the text section. This text reading is always connected with classroom discourse and note-taking about key ideas and how these ideas relate to prior learning, the essential words, and the comprehension canopy question.

Team-Based learning comprehension check. Teachers conduct two comprehension checks comprised of five multiple-choice questions and one open-ended writing question to assess the extent to which students understand key content. Initially, teachers administer the comprehension checks individually to students and the students answer the questions without access to the text or their notes. Then, teachers place students into heterogeneous teams of four or five student and asked them to reach agreement on each answer and cite evidence for each response using their texts and notes. These comprehension checks allow teachers to assess each student's performance, enable students who struggle individually on the comprehension check to learn from their more

knowledgeable peers, and establish procedures for teams to work collaboratively.

Team-based learning knowledge activity. To develop students' ability to think critically about content, explore content from multiple perspectives, and solve problems by applying learned content in a new scenario, PACT includes a team-based learning knowledge application activity for students to work as part of a team to solve a complex problem related to the unit's content. The knowledge application takes place at the end of each unit and serves as a culminating unit activity. For instance, eighth grade students may work on an activity to identify the most important cause of the American Revolution using a game board similar to a sports tournament bracket that includes eight causes. In each round, groups evaluate two causes and identify the more important cause citing evidence from their texts until the team isolates a "champion cause." For all team-based learning knowledge activities, teams conclude the activity by presenting their work to the class and respond to student and teacher questions.

Treatment Fidelity

The research teams put several supports in place to enhance treatment fidelity. For one, all teachers attended 8-10 hours of training and received weekly coaching support. Secondly, teachers received information about the content to teach with all classes (treatment and comparison). Third, teachers received several instructional materials to support successful implementation: (a) semi-scripted PACT lesson plans, (b) a daily schedule identifying the instructional components for each day, (c) student workbooks aligned with the lesson plans, and (d) student materials for the TBL activities. Lastly, the research team was available for support on request from teachers.

Data collection. Vaughn et al. (2015) collected observational data in treatment and

comparison classes. The code-sheets enabled observers to record information about treatment adherence for each of the six PACT instructional elements and overall instructional quality. The research team used the same code-sheet for PACT and typical practice classes. For classes receiving PACT instruction, researchers used the code-sheet to examine the extent to which teachers adhered to the key elements of PACT (treatment adherence) and the overall quality of instruction. For classes receiving typical instruction, researchers used the same code-sheet to measure overall instructional quality and treatment adherence; however, treatment adherence in this case reflects the extent to which PACT instructional techniques crossed over into the comparison condition, as there was no expectation comparison classroom instruction would include the PACT instructional practices.

Vaughn et al. (2015) gathered information using audio recordings. Teachers audio-recorded all 30 lessons for classes assigned to PACT and typical business-as-usual social studies instructional practices. The research team collected information on instruction by coding 10 of the 30 lessons in the 30-lesson sequence. For classes randomly assigned to receive business-as-usual instruction, ten lessons were randomly selected and coded. For classes assigned to PACT, Vaughn et al. randomly selected recordings blocking on intervention components to ensure each intervention component was observed at least twice.

Treatment fidelity coding. The treatment fidelity coding document used in Vaughn et al. (2015) encompassed two dimensions of treatment fidelity: treatment adherence and instructional quality. The code-sheet conceptualized treatment adherence as the extent to which teachers implemented the six essential PACT components. For each

component, coders were offered a checklist of critical procedures and steps of each component and then assigned a rating from a Likert-type scale ranging from 1 (low alignment with instructional procedures) to 4 (high alignment with instructional procedures). For instance, for the knowledge acquisition component, the code-sheet listed the four steps of the knowledge acquisition component: (1) set a purpose for reading, (2) read text section, (3) record notes in learning log, and (4) facilitate short discussion. The coders were trained to evaluate the presence of the steps and then assign a rating. If the component was not expected for that date, then the coders were asked to select "not applicable."

Coders assessed instructional quality using a single, Likert-type item ranging from 1 (low quality) to 5 (high quality) for the entire lesson. During training, coders received a list of quality indicators to help make rating decisions. Examples of quality indicators were "uses instructional time efficiently," "clearly explains concepts," "facilitates active engagement of students during instruction", and "provide frequent feedback to students." The codesheet also included a single, Likert-type item of classroom management, which was also scored from 1 to 5 based on a list of classroom management indicators.

A senior researcher from each site trained two research team members at each site on the fidelity coding procedures and established reliability using the gold standard method (Gwet, 2001). The gold standard method involved coders independently coding a lesson identified by the senior researcher who served as the gold standard. Each coder needed to reach an interrater agreement level of at least 90% with the gold standard's rating. Coders completed additional lessons until they reached this agreement level.

Measures

Trained researcher personnel blind to study condition administered three academic assessments to students at pretest and posttest.

Content knowledge. The Assessment of Social Studies Knowledge and Comprehension (ASK) Content Knowledge subtest was used to assess students' knowledge of the content that was taught by treatment and comparison teachers. The assessment is an untimed measure with 42 four-option, multiple choice items. Items were adopted from Texas previously released Texas and Massachusetts state social studies texts and released advanced placement (AP) tests in social studies developed by the college board. The researcher-developed content measure also contained items to assess concepts and vocabulary not found in the previously released state-developed and AP tests. Alpha coefficient for the ASK Content Knowledge was .89. Additionally, the fit of the underlying confirmatory models with observed responses reported in Vaughn et al. (2013) and Vaughn et al. (2015) suggest the ASK content knowledge measures represent the dimensions they claim to measure (content knowledge and reading comprehension).

Reading comprehension. Reading comprehension was assessed using two measures. The first reading comprehension assessment used was the researcher-developed ASK Reading Comprehension subtest. The test includes 20 four-option, multiple choice items in an untimed format. The assessment contains three expository passage related to, but not explicitly discussed in, the content covered in treatment and comparison classrooms received. The multiple-choice items measured students' ability to decipher main ideas, recognize cause and effect, summarize, and comprehend vocabulary used in context. The ASK Reading Comprehension subtest yielded an alpha coefficient of .85. Students were

also assessed on the Gates MacGinitie Reading Test—4th Edition (GMRT; MacGinitie, MacGinitie, Dreyer, & Hughes, 2000). This timed assessment of reading comprehension requires students to read expository and narrative passages ranging in length from 3 to 15 sentences and answer three to six multiple choice questions per passage. Two alternate test forms were used at pretest and posttest. The internal consistency ranges from .91 to .93 and alternate form reliability is .80 to .87 (MacGinitie et al., 2000)

Analysis Plan

Before addressing the research questions, I investigated the suitability of the data for structural equation modeling (SEM) analyses by evaluating the assumptions that pertain to SEM. Specifically, I examined the skewness and kurtosis levels to assess multivariate normalality. I also evaluated the linearity of the data by inspecting bivariate scatterplots for all variables and seek to identify multivariate outliers using Mahalanobis distance. Additionally, I conducted a confirmatory factor analysis (CFA) to determine whether a latent adherence factor adequately predicted the six PACT instructional components (comprehension canopy, essential words, etc.). This model was assessed using multiple fit indices: chi-square, chi-square, comparative fit index (CFI), Tucker-Lewis index (TLI), standardized root mean square residuals (SRMR), and root mean square error of approximation (RMSEA). I followed established guidelines for evaluating the adequacy of model fit (e.g., Hu & Bentler, 1999).

To address the first research question, I assessed whether a general factor of treatment fidelity predicts instructional quality and instructional adherence. To examine whether treatment adherence and quality predicts content knowledge and reading comprehension performance (RQ2), I used an SEM approach to determine whether and to

what extent treatment adherence and quality predict greater performance in content knowledge and reading comprehension as measured by the ASK Content Knowledge subtest, ASK Comprehension subtest, and the GMRT. I modeled instructional quality as a latent variable to make it possible to examine the interaction term between two latent variables. Single indicator latent variables were specified by fixing the observed measure's factor loading to 1 and fixing its error term to a value 'a' based on the indicator's variance and its reliability (Preacher, Zhang, & Zyphur, 2016). The latent variable 'quality' was created using the one-item measure, "Overall, I consider this teacher's instruction to be…" and the interaction of quality and adherence was modeled as a latent.

Finally, I addressed the third research question by adding teacher's classroom management performance and student characteristics (i.e., pretest performance, limited English proficiency status, special education status, gender) to the existing structural model to determine whether any of these variables moderated the effect of treatment adherence or quality on outcomes. Classroom management and pretest performance were treated as continuous variables. Gender, limited English proficiency (LEP) status, and special education (SPED) status were measured as dichotomous variables: coded 1 for females and 2 for males, coded 0 for non-LEP and 1 for LEP, and 0 for non-SPED and 1 for SPED. For these exploratory analyses, I used the Benjamini-Hochberg correction with the false discovery rate set at 0.05 to adjust for the inflated Type I error involved with the multiple statistical tests (Benjamini & Hochberg, 1995). This procedure helps to control for the small p-values ($p < .05$) that may be due to the increased chance of identifying small p-values when conducting several hypotheses tests.

In all analyses, I accounted for missing data using maximum likelihood estimation

with robust standard errors estimator in Mplus 8.0 (Muthen & Muthen, 2017). I used multilevel modeling with students nested in classes and classes blocked on teacher. There was no significance variance found at the school- and site-levels (intraclass correlations values were below .05), so these levels were not included in the nested model. The outcome variables (i.e., ASK Content Knowledge, ASK Reading Comprehension, GMRT) were collected at the student level. In all structural models, I regressed student pretest scores on the posttest outcome of interest as a student-level covariate.

Chapter 4: Results

Preliminary Analyses

I examined the data for missingness, outliers, non-normality, non-linearity, multicollinearity, and homoscedasticity. Missingness was limited (less than 5% of the total observations for each group) and I assumed this minimal amount of data were missing at random. I accounted for missing data using maximum likelihood estimation with robust standard errors estimator in Mplus 8.0 (Muthen & Muthen, 2017). I identified six univariate outliers (students whose test performance was greater than three standard deviations from the sample mean) and removed these students from the analyses. Additionally, multivariate outliers were inspected using Mahalanobis distances and nine additional cases that appeared to be multivariate outliers were identified. These cases were also removed from the analyses; thus, the final sample included 775 students (for demographic data, see Table 3.1 on pg. 25). Using a cutoff of one standard deviation, descriptive statistics revealed there were no problematic skewness or kurtosis values to suggest univariate or multivariate non-normality except one component of treatment adherence (comprehension canopy) revealed a skewness value of -1.46 (see Table 4.1). Thus, the present study estimated models using maximum likelihood estimation robust standard errors. As shown in Table 4.2, no correlation coefficients between predictor variables exceeded .6, suggesting collinearity problems were not present. Lastly, a visual examination of residual scatterplots showed approximately constant variance, indicating the assumption of homoscedasticity was not violated.

Table 4.1

Means and Standard Deviations of Adherence Components

Components	Mean	Standard Deviation	Range	Skewness	Kurtosis
Comprehension Canopy	3.61	0.47	2-4	-1.46	0.66
Essential Words	3.32	0.52	2-4	-0.06	-0.91
Warm Up	3.24	0.44	2-4	-0.31	-0.78
Knowledge Acquisition	2.82	0.39	2-4	0.15	-0.56
Team-based Learning Knowledge Acquisition	2.57	0.40	2-4	0.67	-0.55

Table 4.2

Correlation Matrix for All Variables

	1	2	3	4	5	6	7	8	9	10	11
1. Comprehension Canopy											
2. Essential Words	**.57**										
3. Warm Up	**.53**	**.42**									
4. Knowledge Acquisition	**.11**	**.46**	**.19**								
5. TBL Knowledge Acquisition	.05	.10	.05	**.59**							
6. Instructional Quality	-.01	-.03	**.11**	**.17**	**.56**						
7. Pretest ASK-CK	-.06	-.01	-.04	**-.09**	.05	**.20**					
8. Posttest ASK-CK	-.02	.02	.07	.02	**.08**	**.22**	**.71**				
9. Pretest ASK-RC	.02	-.03	**-.22**	**-.23**	**-.11**	.06	**.41**	**.42**			
10. Posttest ASK-RC	.00	-.05	**-.23**	**-.20**	**-.06**	**.12**	**.44**	**.50**	**.81**		
11. Pretest GMRT	-.03	-.02	-.05	-.05	**.09**	**.22**	**.70**	**.67**	**.48**	**.51**	
12. Posttest GMRT	-.05	.02	-.01	-.01	**.08**	**.22**	**.67**	**.70**	**.42**	**.47**	**.79**

Note. Bold text indicates a statistically significant correlation ($p < .05$)

I conducted a confirmatory factor analysis (CFA) with six indicators of treatment adherence: comprehension canopy, essential words, warm up, knowledge acquisition, TBL knowledge acquisition, and TBL comprehension check. Models with TBL comprehension check were less reliably aligned with the observed data. Moreover, bivariate correlations between the components and other components of adherence were not different from zero, suggesting the component may not have been a dimension of adherence. Figure 4.1 presents the measurement model for treatment adherence. The model statistics for the included adherence components are shown in Table 4.3. The five measured treatment adherence indicators were all positively and significantly associated to the latent adherence variable. To determine model fit, I used the cut-off values put forth by Hu and Bentler (1999): chi-square p-value = .05, RMSEA = .06, CFI = .90, TLI = .90, SRMR = .08. With the exception of the TLI value, the measurement model results indicated adequate model fit according to the widely-used fit indices with χ^2 (5)= 11.96, *p*=.04*;* RMSEA = .04; CFI=.90; TLI=.80; SRMR=.06. ≥

Figure 4.1

Measurement Model for Treatment Adherence.

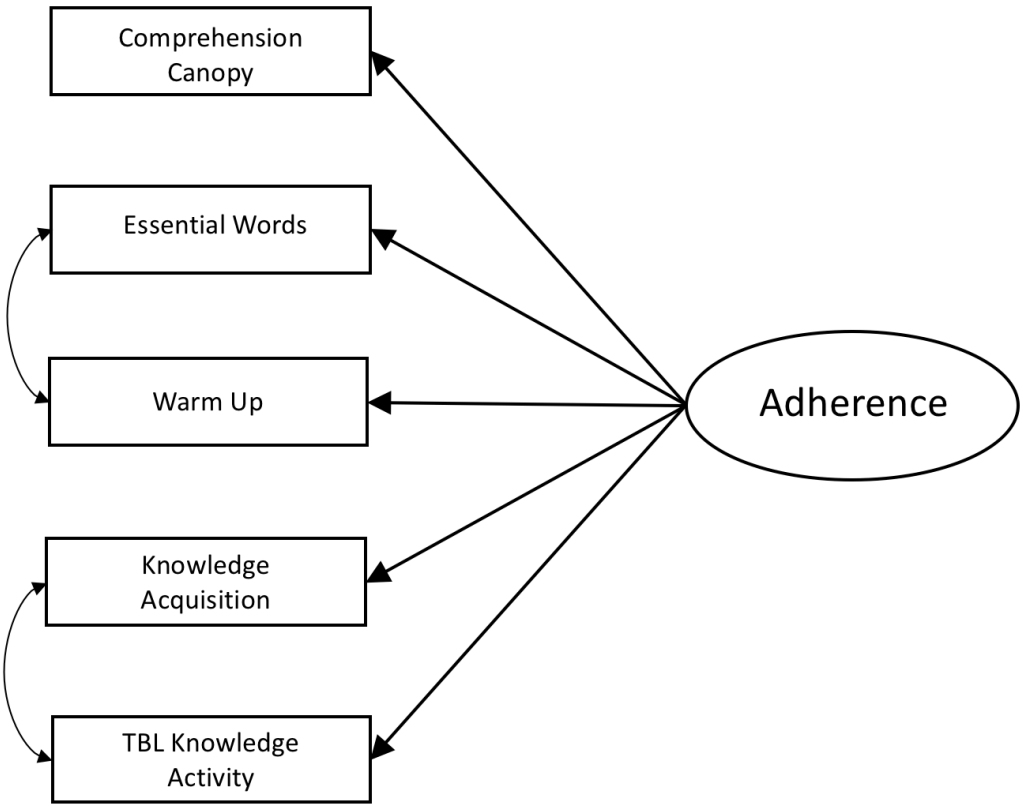


Table 4.3

Adherence Measurement Model Coefficients, Residuals, R^2, And Intercepts

Components	Loading (SE)	Residual (SE)	R^2	Intercept (SE)
Comprehension Canopy	.70 (.13)	0.51(.18)	.49 (.18)	5.25 (.79)
Essential Words	.80 (.03)	.36(.06)	.63(.06)	4.49 (.46)
Warm Up	.75 (.05)	.44 (.07)	.56 (.07)	4.80 (.46)
Knowledge Acquisition	.39 (.17)	.85 (.13)	.15 (.13)	4.51 (.42)
Team-based Learning Knowledge Acquisition	.10 (.17)	.99 (.03)	.01 (.03)	4.03 (.34)

Notes. Fit indices χ^2 (5)= 11.96, p=.04; RMSEA = .04; CFI=.90; TLI=.80; SRMR=.06.

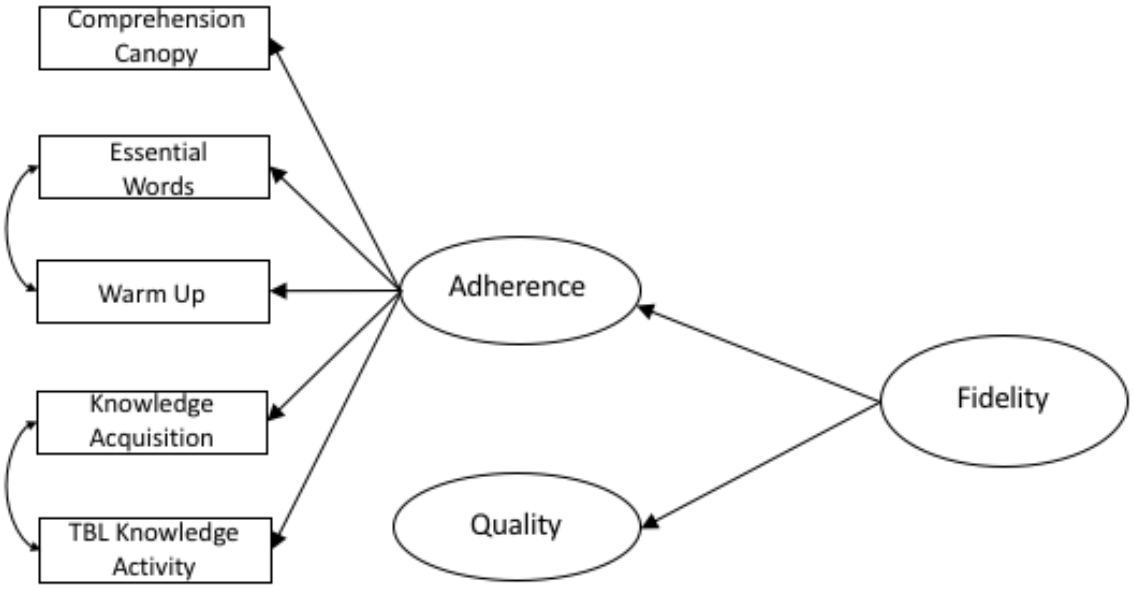
Results from Main Data Analyses

Research Question 1

I examined whether instructional adherence and quality were predicted by a single factor of treatment fidelity using confirmatory factor analysis. As shown in Figure 4.2, the hypothesized measurement model tested was nominally just-identified based on the model structure, indicating the model had zero degrees of freedom. However, results from the confirmatory factor analysis revealed the latent fidelity measurement model was empirically under-identified based on the sample data analyzed. Empirical under-identification, in this case, inhibited the confirmatory factor analysis from reaching a final solution (i.e., producing a valid set of parameter estimates). Empirically under-identified solutions can occur when a second-order factor has a near-zero effect on the first-order factor in conceptually just-identified models (Brown, 2006; Kenny, 1979; Rindskopf, 1984). In this study, the latent factor of treatment fidelity (second order factor) did not have a direct effect on treatment adherence when instructional quality was fixed to one as the marker variable and treatment adherence was free to vary. Equally, the latent factor of fidelity did not have a direct effect on instructional quality when it was free to vary and treatment adherence was set as the marker variable. It is likely that empirical under-identification was a function of having only two first-order factors (quality and adherence) in the model and that there was only a weak correlation ($r = .18$) between these factors. Nonetheless, this result indicates that treatment adherence and quality in this sample are distinct factors.

Figure 4.2

Hypothesized Model for Treatment Fidelity



Research Question 2

To examine the extent to which adherence and quality predicted performance in content knowledge and reading comprehension, the academic outcomes (ASK Content Knowledge, ASK Reading Comprehension, GMRT) were regressed on the latent variables of adherence and quality as class-level predictors and pretest performance as a student-level covariate. The structural model is shown in Figure 4.3 and the standardized coefficients are presented in Table 4.4.

Figure 4.3

Structural Model for Treatment Adherence and Quality on Student Outcomes

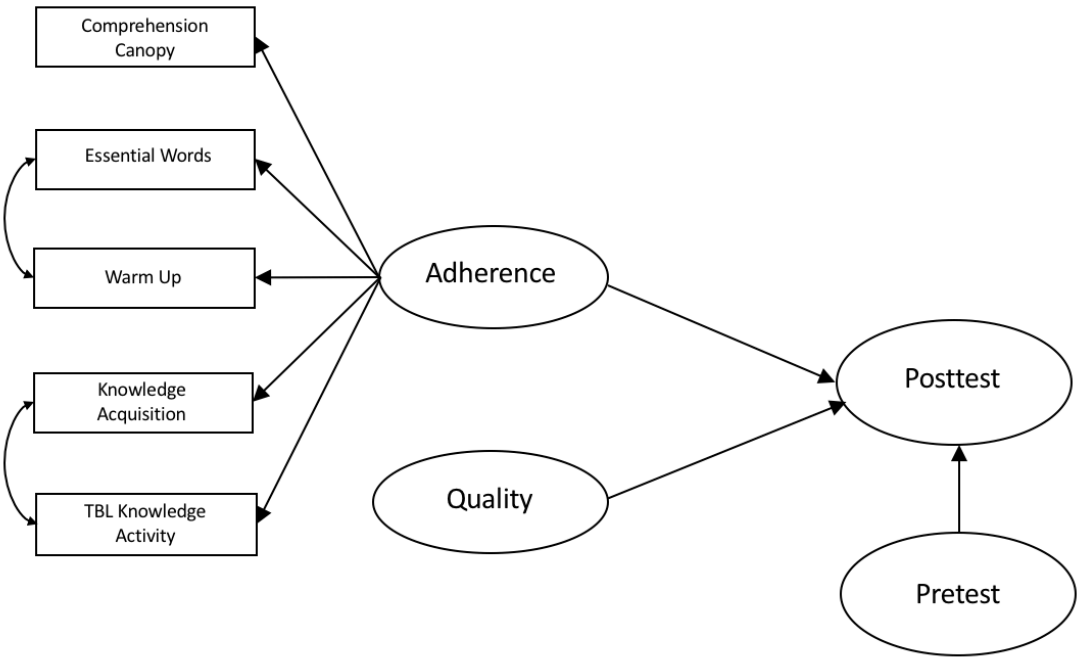


Table 4.4

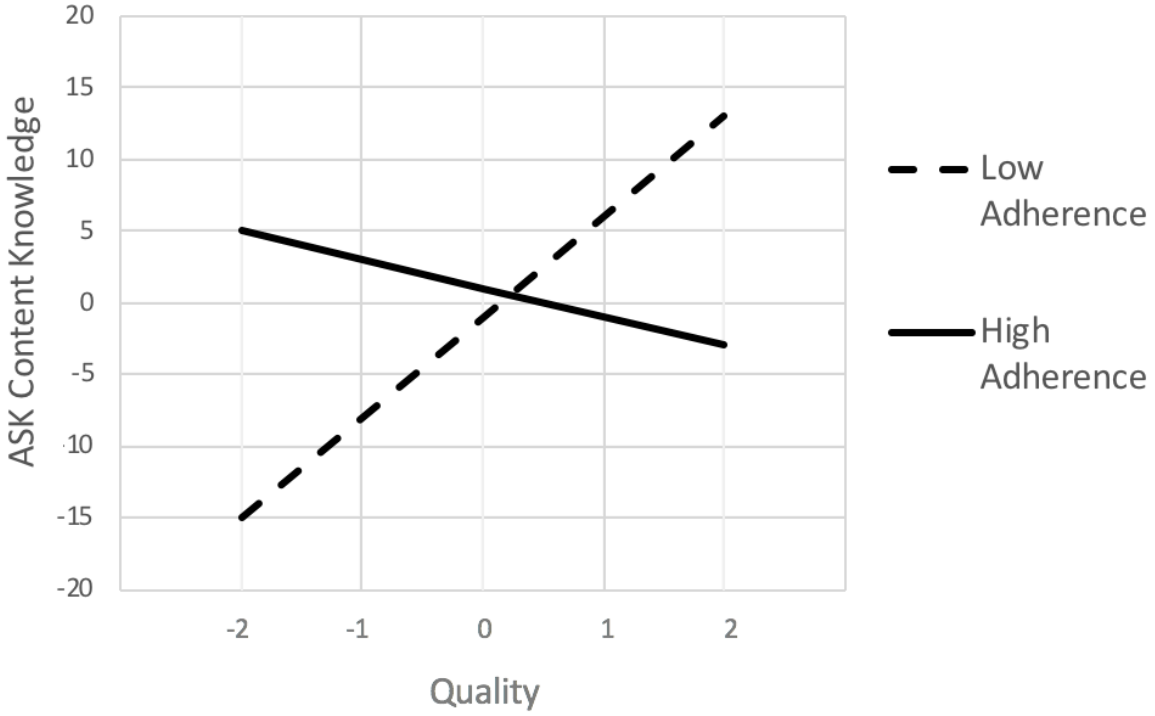
Parameter Estimates and Significance Levels for Treatment Adherence and Quality on Content Knowledge and Reading Comprehension Outcomes

Parameter	Estimate	S.E.	Est./S.E.	*p*-value*
ASK-CK ON				
ASK $CK_{pretest}$	.72	.01	52.86	.000
Instructional Adherence	.10	.05	2.42	.02
Instructional Quality	.14	.05	2.72	.01
Quality * Adherence	-.26	.07	-3.67	.00
ASK-RC ON				
ASK-RC $_{pretest}$	.76	.01	66.03	.00
Instructional Adherence	.00	.04	.02	.98
Instructional Quality	.10	.03	3.05	.00
Quality * Adherence	-.06	.05	-1.22	.22
GMRT ON				
$Gmrt_{pretest}$	.79	.01	60.80	.00
Instructional Adherence	.04	.05	.94	.34
Instructional Quality	.06	.04	1.52	.13
Quality * Adherence	-.09	.08	-1.045	.30

Content Knowledge. On the ASK Content Knowledge subtest, both adherence and quality were significantly positive predictors of posttest performance, suggesting that increases in both dimensions of treatment fidelity were associated with improvements in posttest performance. However, a significant interaction between treatment adherence and quality was present. As shown in Figure 4.4, this interaction term indicates that greater instructional quality was associated with higher levels of posttest performance for students in classrooms with relatively low levels of instructional adherence. It is important to interpret these findings considering that treatment adherence was on average high (M = 3.08, SD = .48 on a scale from 1 to 4) whereas instructional quality scores were not, on average, as high (M = 2.82, SD = .71 on a scale from 1 to 5). Thus, references to low levels of treatment adherence should be interpreted as relatively lower adherence scores that actually reflect moderate rather than low instructional adherence. However, for students in classrooms with teachers with high adherence, higher levels of instructional quality were not positively associated with posttest performance. Results actually showed a weak negative association between instructional quality and posttest performance for students in classrooms with high treatment adherence. Also, results suggested less variance in student outcomes when adherence was high than when it was low. Yet, it appears low adherence paired with high-quality instruction leads to greater posttest performance than high adherence and low-quality instruction (see Figure 4.4).

Figure 4.4

Interaction Effect of Instructional Quality and Adherence on Posttest ASK Content Knowledge Performance



Reading Comprehension. Results on the ASK Reading Comprehension subtest indicated instructional quality was a significant, positive predictor of posttest reading comprehension performance. Specifically, a one-point increase in instructional quality was associated with a .1 standard deviation increase on the ASK reading comprehension subtest. Treatment adherence was not a significant predictor and no significant interaction between instructional quality and treatment adherence was present. Results on the GMRT found instructional adherence and quality were not predictive of posttest performance, nor was the interaction between these variables significant.

Research Question 3.

The final research question addressed the extent to which teachers' classroom management and student-level characteristics (pretest performance on outcome measures, limited-English proficiency status, special education status, and gender) moderated the effects of treatment adherence and quality on student outcomes. Standardized coefficients and significance levels are presented for the teacher- and student-level moderators of instructional adherence and quality in Table 4.5 and 4.6, respectively. The results are presented below by moderator.

Table 4.5

Parameter Estimates and Significance Levels for Moderators of Treatment Adherence on Content Knowledge and Reading Comprehension Outcomes

Parameter	Estimate	S.E.	Est./S.E.	*p*-value
ASK-CK ON				
Instructional Adherence	.11	.05	2.42	.02
ASK $CK_{pretest}$	.87	.02	39.91	.00
Classroom Management	-.10	.06	-1.56	.12
Limited-English Proficient	-.06	.02	-3.70	.00
Gender	-.02	.02	1.10	.27
Special Education	-.10	.01	-5.32	.00
Adherence * ASK $CK_{pretest}$	-.39	.07	-5.85	.00[a]
Adherence * Classroom Management	.16	.06	2.80	.01[a]
Adherence * Limited-English Proficient	.02	.02	1.11	.27
Adherence * Gender	.00	.03	-.16	.87
Adherence * Special Education	.00	.03	.12	.63
ASK-RC ON				
Instructional Adherence	.00	.04	.02	.98
ASK $RC_{pretest}$	.76	.01	66.03	.00
Classroom Management	.02	.03	.86	.39
Limited-English Proficient	-.19	.05	-3.61	.00
Gender	.02	.02	1.19	.23
Special Education	-.05	.02	-2.91	.00
Adherence * ASK $RC_{pretest}$	-.03	.02	-1.16	.24
Adherence * Classroom Management	.03	.04	.60	.55
Adherence * Limited-English Proficient	.34	.19	1.81	.07
Adherence * Gender	-.01	.03	-.40	.69
Adherence * Special Education	-.01	.02	-.29	.77
GMRT ON				
Instructional Adherence	.04	.05	.94	.34
$GMRT_{pretest}$	.79	.01	60.80	.00
Classroom Management	-.07	.06	-1.22	.22
Limited-English Proficient	-.05	.02	-2.88	.00
Gender	.02	.02	.97	.33
Special Education	-.04	.02	-1.94	.05
Adherence * $GMRT_{pretest}$	-.00	.00	-1.90	.06
Adherence * Classroom Management	.01	.04	.32	.75
Adherence * Limited-English Proficient	-.02	.02	.81	.42
Adherence * Gender	-.03	.03	-1.15	.25
Adherence * Special Education	.01	.02	.60	.55

Note. ASK-CK – ASK Content Knowledge; ASK-RC – ASK Reading Comprehension; GMRT – Gates-MacGinitie Reading Test-4th edition. [a] $p < .05$ after Benjamini-Hochberg correction.

Table 4.6

Parameter Estimates and Significance Levels for Moderators of Instructional Quality on Content Knowledge and Reading Comprehension Outcomes

Parameter	Estimate	S.E.	Est./S.E.	*p*-value*
ASK-CK ON				
Instructional Quality	1.24	.48	2.57	.01
ASK Content Knowledge$_{pretest}$	.74	.01	58	.00
Classroom Management	-.04	.07	-.55	.58
Limited-English Proficient	-.06	.02	-3.70	.00
Gender	.02	.02	1.10	.27
Special Education	-.10	.02	-5.32	.00
Quality * ASK CK$_{pretest}$	-.15	.04	-3.75	.00[a]
Quality * Classroom Management	.16	.06	2.76	.01[a]
Quality * Limited-English Proficient	.02	.02	1.11	.27
Quality * Gender	-.00	.03	-.16	.87
Quality * Special Education	.00	.03	.12	.90
ASK-RC ON				
Instructional quality	.10	.03	3.05	.00
ASK Reading Comprehension$_{pretest}$	.75	.02	43.28	.00
Classroom Management	-.05	.05	-.98	.33
Limited-English Proficient	-.17	.07	-2.58	.01
Gender	-.06	.09	-.62	.53
Special Education	-.02	.08	-.21	.84
Quality * ASK RC$_{pretest}$	.03	.04	.79	.43
Quality * Classroom Management	.02	.03	.61	.54
Quality * Limited-English Proficient	.10	.08	1.30	.20
Quality * Gender	.07	.12	.55	.58
Quality * Special Education	-.01	.08	-.152	.88
GMRT ON				
Instructional Quality	.06	.04	1.52	.13
GMRT$_{pretest}$	.80	.01	82.84	.00
Classroom Management	-.09	.12	-1.22	.22
Limited-English Proficient	-.03	.09	-.27	.79
Gender	02	.10	.17	.86
Special Education	.03	.11	.29	.77
Quality * GMRT$_{pretest}$	.05	.04	1.39	.17
Quality * Classroom Management	-.03	.07	-.38	.71
Quality * Limited-English Proficient	-.02	.08	-.18	.86
Quality * Gender	.01	.14	.17	.86
Quality * Special Education	-.06	.09	-.69	.49

Note. ASK-CK – ASK Content Knowledge; ASK-RC – ASK Reading Comprehension; GMRT – Gates-MacGinitie Reading Test-4th edition. [a] $p < .05$ after Benjamini-Hochberg correction.

Classroom Management. I investigated one teacher-level moderator: classroom management. Results showed classroom management did not significantly predict posttest performance on outcome measures when added to the structural model with instructional adherence, quality, and pretest performance. Although there was no direct effect of classroom management on posttest outcomes, classroom management did moderate the effects of instructional adherence and quality on ASK content knowledge. As shown in Figures 4.5 and 4.6, for both treatment adherence and quality, the strength of the relations between treatment fidelity and content knowledge was weak for students in classrooms with low classroom management. This suggests the positive effects of high instructional quality and adherence on students' content knowledge were not present for students in classrooms with low classroom management. For students in classrooms with high classroom management, the relation between treatment fidelity and comprehension knowledge performance was apparent: increases in treatment fidelity were associated with improved student outcomes. Moderation effects were not present on the secondary outcome of reading comprehension.

Figure 4.5
Interaction Effect of Instructional Quality and Classroom Management on Posttest ASK Content Knowledge Performance

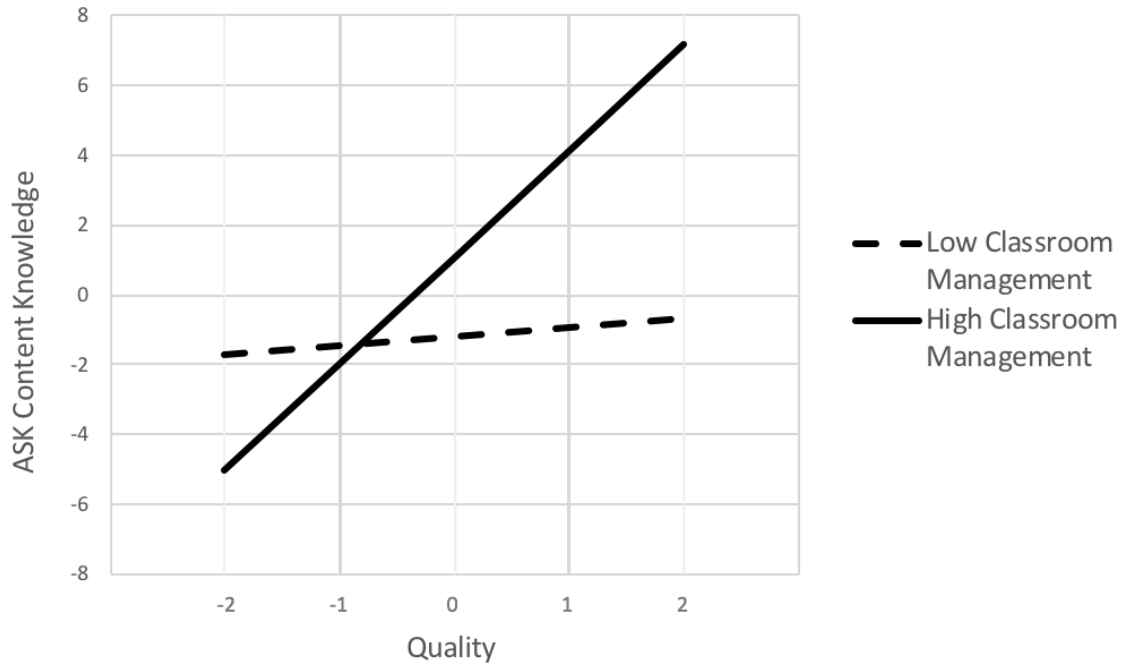
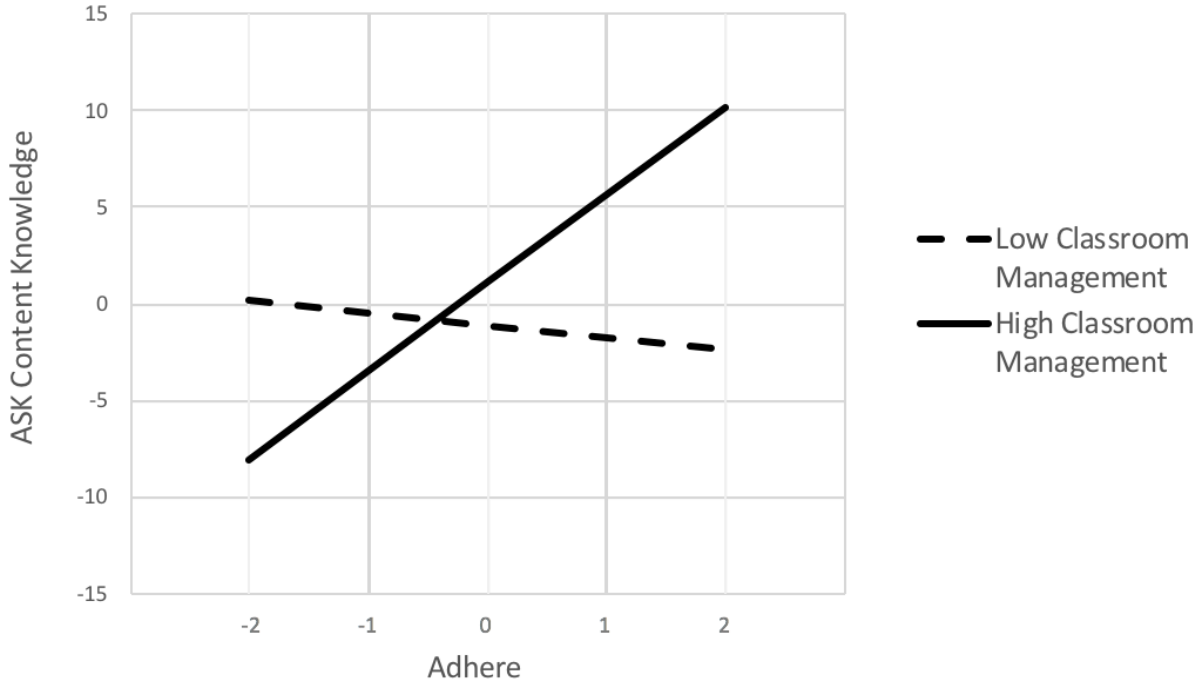


Figure 4.6
Interaction Effect of Instructional Adherence and Classroom Management on Posttest ASK Content Knowledge Performance



Pretest Performance on Academic Outcomes. The next moderator examined was pretest performance. As expected, pretest performance was a strong predictor of content knowledge and reading comprehension outcomes (standardized coefficients exceeded .7 on all measures). Results showed pretest performance on ASK content acquisition moderated the effects of instructional quality and adherence on ASK posttest performance. As displayed in Figures 4.7 and 4.8, higher levels of instructional adherence and quality were associated with greater posttest ASK content knowledge scores for students with low pretest scores. For students with high initial performance on the ASK content measure, the effects of treatment adherence and quality and posttest performance varied slightly based on the dimension of fidelity (quality or adherence) examined. For students with high initial performance on the ASK content knowledge, higher levels of instructional quality led to similar levels of posttest performance, indicating that instructional quality was not associated with ASK content knowledge outcomes for student with high initial performance. However, for higher performing students, higher levels of instructional adherence do appear to be associated with higher levels of posttest performance; however, this relation does not appear to be as strong as it is for lower performing students. The finding that pretest performance moderated the effects of treatment adherence and quality on posttest outcomes was not replicated on the secondary outcome of reading comprehension.

Figure 4.7

Interaction Effect of Instructional Quality and Pretest Content Knowledge on Posttest ASK Content Knowledge Performance

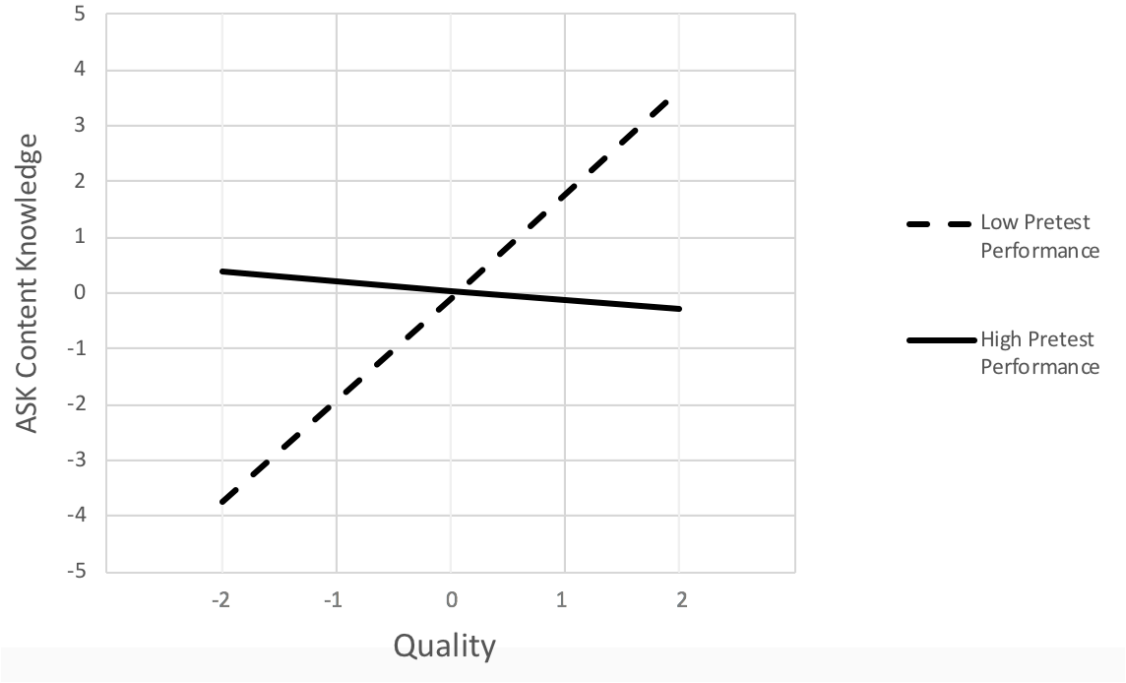
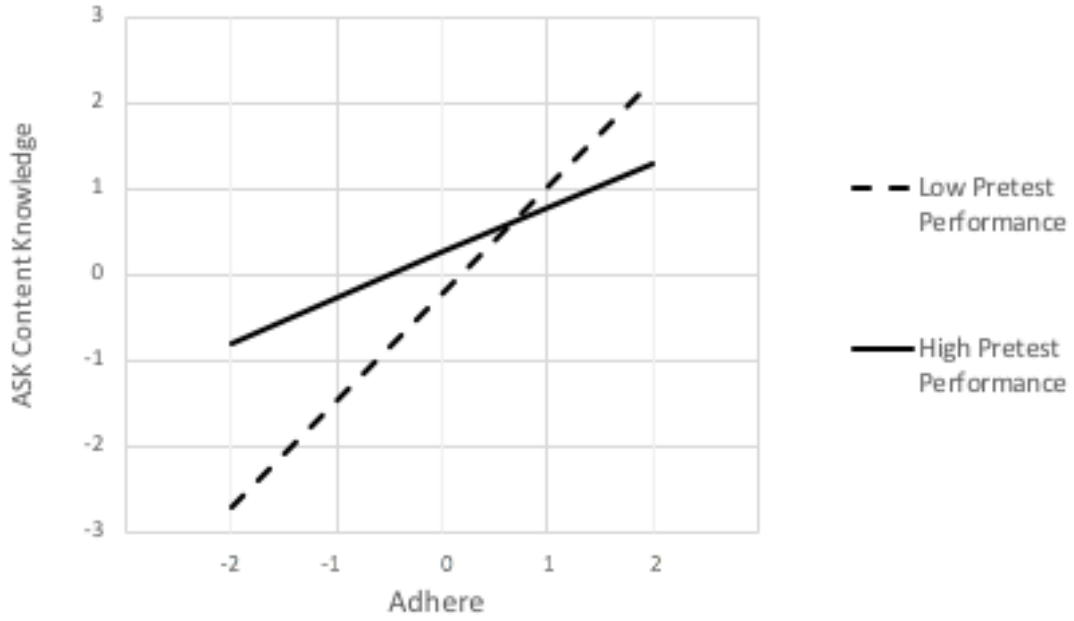


Figure 4.8

Interaction Effect of Instructional Adherence and Pretest Content Knowledge on Posttest ASK Content Knowledge Performance



Other Student-level Moderators. I examined three additional student-level moderators: limited-English proficiency status, special education status, and gender. None of these categorical variables significantly moderated the effects of treatment adherence or quality on student outcomes.

Chapter 5: Discussion

Treatment fidelity is an important methodological consideration and data source for educational researchers aiming to better understand the conditions under which instructional practices are efficacious. The purpose of this study was to investigate the influences of treatment adherence and quality on academic outcomes for eighth grade students who received a researcher-designed and classroom teacher implemented content-area reading program. I was interested in (a) understanding whether treatment adherence and quality are manifestations of a broader treatment fidelity construct (RQ1), (b) examining the extent to which treatment adherence and quality account for variance in student content knowledge and reading comprehension outcomes for eighth graders whose teachers implemented the PACT content-area reading practices (RQ2), and (c) exploring potential teacher- and student-level characteristics that may moderate the effects of treatment adherence and quality on student outcomes (RQ3).

Instructional Adherence and Quality as Dimensions of Treatment Fidelity

The first research question tested the hypothesis that instructional adherence and quality contribute to a general factor of treatment fidelity. Confirmatory model results indicated that treatment adherence and quality were distinct constructs that were not adequately predicted by a general factor. This finding aligns with the results of Guo et al. (2016). Reanalyzing the results of a literacy intervention for preschool students, Guo et al. (2016) found three dimensions of treatment fidelity (adherence/dosage, treatment differentiation, and student engagement) were best modeled as independent factors. However, these findings do not align with the results reported by Fogarty and colleagues (2014). Collecting treatment fidelity information on treatment adherence, quality, dosage,

student engagement, and differentiation, factor analyses by Fogarty et al. revealed all five dimensions were significantly related to a general factor of fidelity. Crucially, although there was overlap between the dimensions of treatment fidelity investigated in the present and preceding studies, there were differences in the number and type of treatment fidelity dimensions investigated in each study. Moreover, the studies employed different data collection methods. Guo et al., for instance, used teachers' submission of weekly reports to assess treatment adherence whereas Fogarty et al. and the current study used observational data to measure treatment adherence. Another complicating difference between the studies is that the nature of the instructional programs and participants varied substantially across the studies. The Fogarty et al study was conducted with students with significant reading difficulties as an intensive intervention whereas the data utilized in the PACT study was from students representing a range of reading abilities in a typical classroom. Given these mixed findings, it may be there are contextual conditions that influence the role of fidelity as it relates to programs, participants, and conceptualization.

Effects of PACT Treatment Adherence and Quality on Student Outcomes

The second research question examined whether treatment adherence and quality predicted student outcomes for students who received the PACT content-area reading practices. This research expanded on the work of Vaughn et al. (2015). Their original study evaluated data for the full sample (treatment and control) to find significant differences between treatment and comparison on the ASK Content Knowledge measure. Vaughn et al. also found that treatment adherence fully mediated this effect, which strengthens the causal relation between PACT and improved content knowledge among treatment students. However, Vaughn et al. (2015) did not detect significant effects from PACT on the

secondary outcome of reading comprehension, nor did the authors find treatment adherence mediated the relation between treatment condition and reading comprehension performance.

This study extended Vaughn et al.'s previous work on treatment fidelity by examining the extent to which treatment adherence and quality predict variation among students randomized to the PACT treatment. The present study reported treatment adherence and quality were significant positive predictors of content knowledge. Importantly, results showed treatment adherence and quality interacted with each other. Specifically, instructional quality was positively associated with content knowledge outcomes when treatment adherence was low, but not when treatment adherence was high. This finding suggests that instructional quality is particularly important when teachers have relatively low adherence to the PACT practices; however, instructional quality is less critical when adherence is high.

The treatment fidelity findings from the content knowledge outcome measure were not replicated on the secondary outcome of reading comprehension. Understanding the goals of the PACT instructional practices may help to explain why there was a different pattern of results for reading comprehension than those present for content knowledge. Vaughn and colleagues (2013) developed the PACT instructional practices to aid middle-school history teachers in meeting their primary goal—supporting student acquisition of history content. Considering reading comprehension was a secondary study outcome, it may not be surprising that treatment adherence and quality for a program primarily targeting content knowledge acquisition were not consistently associated with reading comprehension outcomes.

Although treatment fidelity dimensions were not predictive of reading comprehension outcomes, findings on the primary outcome align with previous research establishing multiple dimensions of treatment fidelity—including treatment adherence and quality—were related to student outcomes (Fogarty et al., 2014; Wolgemuth et al., 2014). Wolgemuth and colleagues (2014), for instance, reported treatment adherence and instructional quality were related to the targeted literacy outcomes for students in the early elementary grades who received a reading intervention. It is worth noting counter-examples exist (Benner, Nelson, Stage, and Ralston, 2011; Hamre et al., 2010; Mendive, Weiland, Yoshikawa, & Snow, 2015). Some studies testing multiple dimensions of treatment fidelity and their relations to student outcomes identified only some dimensions of treatment fidelity were predictive of variation in student outcomes. For example, Benner, Nelson, Stage, and Ralston (2011) reported the relation between treatment adherence and quality and reading outcomes in a middle-school reading intervention study. Benner et al. reported treatment adherence items were consistently associated with student outcomes, however instructional quality items were not. Conversely, in the context of an early childhood language and literacy study, Hamre et al. (2010) reported instructional quality was strongly associated with student outcomes, but adherence and dosage were not.

The present study advances knowledge about treatment fidelity by examining whether dimensions of treatment fidelity interact with each other. To my knowledge, this is the first study to investigate interaction effects of treatment fidelity dimensions on student outcomes. Although there is not previous work to help explain the interaction between treatment adherence and quality, it stands to reason that adherence and quality may vary for a given teacher and that their effects may be conditional on levels of the other variable.

Why might we expect a teacher to demonstrate varying levels of instructional quality and adherence? Let us consider the circumstance of the PACT RCT. It is likely that PACT represented a shift in the teacher's instructional approach. This notion is supported by the Vaughn et al.'s (2015) findings showing considerable differentiation between the instructional practices teachers used during PACT and their BAU classes. We might expect there to be a trade-off between instructional quality and adherence for teachers implementing a set of instructional practices for the first time. For instance, a teacher primarily focused on adhering to the new set of PACT practices may fail to maintain appropriate lesson pace, facilitate active engagement, and monitor student learning, which would reflect lower levels of instructional quality. Conversely, a teacher committed to maintaining quality instruction may focus on these facets of quality instruction at the expense of adhering to the exact procedures of PACT instruction.

Exploring Potential Moderators of Treatment Fidelity

This study tested five potential moderators of the effects of treatment adherence and quality on student outcomes: teacher's classroom management proficiency, student's pretest performance, special education status, limited English proficiency status, and gender. A previous study that found student characteristics influenced the association between treatment fidelity and academic outcomes at posttest (Hamre et al., 2010). Particularly, Hamre et al. (2010) found the association between instructional quality and student outcomes was stronger for students with low initial preliteracy skills and whose home language was not English. These findings, along with other research observing student by treatment interactions (e.g., Connor, Morrison, & Petrella, 2004), provided the motivation for exploring potential moderators of treatment fidelity.

Compatible with the findings of Hamre et al., the reported findings in this paper suggest there are factors that may moderate the effects of treatment adherence and quality on student outcomes. Classroom management and student pretest performance both moderated the effects of treatment adherence and quality on the content acquisition. For classroom management, results showed treatment adherence was positively associated with content acquisition for students in classes with high classroom management; however, there was little to no association between treatment adherence and quality for students in classes with low levels of classroom management. Thus, poor classroom management appears to neutralize the positive effects of higher levels of treatment fidelity on student outcomes. This finding aligns with previous research suggesting effective classroom management practices are positively associated with student outcomes (e.g., Korpershoek, Harms, de Boer, van Kuijk, & Doolaard, 2016).

Student pretest performance also moderated the effects of treatment adherence and quality on the primary outcome of content knowledge. Similar to Hamre et al., the reported findings showed the associations between treatment fidelity, operationalized in this paper as adherence and quality, and student outcomes were stronger for lower performing students. This pattern of findings underscores the importance of instructional quality and adherence to treatment for low performing students.

Limitations and Future Research

There are several considerations when interpreting the findings from this study. Although the student sample under study was ample (n =775), the study would have benefited from a larger sample of participating teachers (19) and classrooms (47). There are a few potential consequences from the limited sample of teachers and classrooms including

a heightened risk for Type II errors and limited confidence findings will generalize to other populations. Sample size requirements for structural equation modeling vary based on the number of factors and the magnitude of their factor loadings. In this study, the limited sample size may have particularly influenced whether a general factor of treatment fidelity predicted adherence and quality, as these factors were estimated at the class-level. Also related to sample size, although the student demographics resembled U.S. demographic figures for school-age children in terms of racial/ethnic makeup and socioeconomic status, the limited number of limited English proficient students and special education students may have affected the results.

There are also a few limitations related to the treatment fidelity data used in this study. This study examined only two of the five aspects of treatment fidelity borne out of theoretical work (Dane & Schneider, 1998) and tested empirically (Fogarty et al., 2014). Future research examining dosage, participant responsiveness, and program differentiation alongside treatment adherence and quality may help to provide further information about the conditions under which PACT and other instructional programs are efficacious.

Another limitation of this study was the use of a single indicator of instructional quality. Collecting data and analyzing multiple indicators of instructional quality (e.g., feedback quality, pacing, providing ample opportunities to respond) may help pinpoint the specific elements of instructional quality that account for variations in student outcomes. Furthermore, this information would hold instructive value to those who are interested in enhancing teacher quality in future implementations of the PACT program. The current study found treatment fidelity and quality scores were generally high and the range of scores was limited, which tends to diminish correlation coefficients. This may not be a

study limitation as much as a consequence of studying treatment fidelity when instructional implementers are highly-qualified teachers provided with 8-10 hours of training, highly-specified instructional materials, and ongoing coaching support. However, future studies may consider alternative approaches to measuring treatment adherence and quality to enhance the reliability of fidelity estimates, as previous studies have shown small changes to Likert-type scales, such as adding one additional scale item can lead to significant improvements in reliability estimates (Adelson & McCoach, 2010). Finally, it is important to note treatment fidelity, quality, and classroom management were scored by the same observer and, therefore, may have been susceptible to halo effects.

Implications

The findings from this study contain implications for practice and future research. For practitioners, this study sheds light on the implementation conditions under which the PACT practices are associated with enhanced outcomes. Study findings showed that treatment adherence and instructional quality are important when teachers are implementing PACT for the first time. Moreover, study findings indicate instructional adherence and quality are particularly important for students who demonstrate lower achievement at the start of implementation. Study conclusions are complicated by evidence showing the effects of treatment adherence and quality depend on each other; however, the interplay between adherence and quality may have important consequences for those intending to implement PACT in the future. The results showed instructional quality was increasingly predictive of outcomes when adherence was relatively low; however, the instructional quality did not appear to be associated with outcomes when teachers adhered closely to the PACT components. This suggests that adhering closely to the PACT

practices may be particularly important for teachers who struggle to provide high-quality instruction. Study findings also suggest teachers and administrators who are considering implementing PACT in the future would be prudent to consider coupling PACT with evidence-based classroom management techniques. Results showed that the positive effects of treatment adherence and quality on content knowledge acquisition were limited by low levels of classroom management.

The findings of this study also reveal several opportunities for future research. First, future research on treatment fidelity may benefit from seeking to replicate the findings of this study in future implementation of PACT or other instructional programs. I interpret the significant interaction between treatment adherence and quality on student outcomes and the moderating role of classroom management on the effects of treatment fidelity as novel findings that require replication to assure the results are valid and reliable. Second, this study suggests that it may be useful for future examinations of treatment adherence and quality to examine the interaction between these variables when implementers are teaching a program for the first time. Although this study was limited to examining two dimensions of treatment fidelity, there may be reasons to think other dimensions of treatment fidelity (e.g., dosage, student responsiveness, and differentiation) may also interact with each other. Third, research examining teacher- and student-level characteristics that may influence the effects of treatment fidelity may also enhance our understanding of how and when treatment fidelity impacts student gains. The present study found teacher's classroom management was not directly related to student outcomes; however, it did influence the associations between treatment adherence and quality and student outcomes. Future research examining classroom management and other teacher-

level characteristics, such as teacher's perceptions of a program's efficacy, may identify other factors that influence the effects of fidelity on student outcomes.

Identifying evidence-based instructional practices and the conditions under which these practices enhance student outcomes is a central purpose of educational research. In-depth examinations of treatment fidelity support the accurate interpretations of study results and can help to inform decisions about scaling up interventions. Although researchers have long considered treatment fidelity data vital to experimental validity, reviews of treatment fidelity indicate intervention researchers frequently fail to report treatment fidelity information in their published reports (e.g., O'Donnell, 2008). However, a recent review of treatment fidelity reporting suggests the frequency of fidelity reporting has increased in recent years (Capin et al., 2017). Importantly, intervention researchers are beginning to explore treatment fidelity in sophisticated ways: (1) examining multiple dimensions of treatment fidelity, and (2) assessing the relation between treatment fidelity dimensions and student outcomes. The findings from this study provide additional empirical evidence that examining multiple dimensions of treatment fidelity will help to explain differences in student outcomes.

References

Adelson, J. L., & McCoach, D. B. (2010). Measuring the mathematical attitudes of elementary students: The effects of a 4-point or 5-point Likert-type scale. *Educational and Psychological measurement*, *70*(5), 796-807.

Anders, P. L. (2002). Secondary reading programs: A story of what was. In D. L. Schallert, C. M. Fairbanks, J. Worthy, B. Maloch, & J. V. Hoffman (Eds.), *51st Yearbook of the National Reading Conference* (pp. 82-93).

Anders, P. L., Hoffman, J. V., & Duffy, G. G. (2000). Teaching teachers to teach reading: Paradigm shifts, persistent problems, and challenges. *Handbook of Reading Research*, *3*, 719-742.

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., ... & Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology, 23*(5), 443.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300

Benner, G. J., Nelson, J. R., Stage, S. A., & Ralston, N. C. (2011). The influence of fidelity of implementation on the reading outcomes of middle school students experiencing reading difficulties. *Remedial and Special Education*, *32*(1), 79-88.

Berkeley, S., Scruggs, T. E., & Mastropieri, M. A. (2010). Reading comprehension instruction for students with learning disabilities, 1995—2006: A Meta-analysis. *Remedial and Special Education*, *31*(6), 423-436.

Berman, R. A., & Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse processes*, *43*(2), 79-120.

Biancarosa, G. & Snow, C.E. (2004). Reading next: A vision for action and research in middle and high school literacy. A report to the Carnegie Corporation of New York. Washington, DC: Alliance for Excellent Education.

Buckley, P., Moore, B., Boardman, A. G., Arya, D. J., & Maul, A. (2017). Validating a Fidelity Scale to Understand Intervention Effects in Classroom-Based Studies. American Educational Research Journal, 54(6), 1378-1413.

Boardman, A. G., Klingner, J. K., Buckley, P., Annamma, S., & Lasser, C. J. (2015). The efficacy of Collaborative Strategic Reading in middle school science and social studies classes. *Reading and Writing*, *28*(9), 1257-1283.

Borrelli, B., Sepinwall, D., Ernst, D., Bellg, A. J., Czajkowski, S., Breger, R., ... & Resnick, B. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology, 73*(5), 852.

Breitenstein, S. M., Fogg, L., Garvey, C., Hill, C., Resnick, B., & Gross, D. (2010). Measuring implementation fidelity in a community-based parenting intervention. *Nursing Research*, *59*(3), 158.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York: Guilford Publications.

Byrne, B.M. (2010). *Structural equation modeling with Mplus: Basic concepts, applications, and programming* (2nd ed.). New York, NY: Routledge.

Capin, P., Walker, M. A., Vaughn, S., & Wanzek, J. (2017). Examining How Treatment Fidelity Is Supported, Measured, and Reported in K–3 Reading Intervention Research. *Educational Psychology Review*, 1-35

Carnegie Council on Advancing Adolescent Literacy. (2010). *Time to act: An agenda for advancing adolescent literacy for college and career success*. New York, NY: Carnegie Corporation of New York.

Cheung, A., Slavin, R. E., Kim, E., & Lake, C. (2017). Effective secondary science programs: A best-evidence synthesis. *Journal of Research in Science Teaching*, *54*(1), 58-81.

Connor, C. M., Morrison, F. J., & Petrella, J. N. (2004). Effective Reading Comprehension Instruction: Examining Child x Instruction Interactions. *Journal of Educational Psychology*, *96*(4), 682.

Cook, T. D., Campbell, D. T., & Day, A. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control?. *Clinical Psychology Review, 18*(1), 23-45.

De Oliveira, L.C. (2010). Nouns in history: Packing information, expanding explanations, and structuring reasoning. The History Teacher, 43(2), 191–203.

Domitrovich, C. E., Gest, S. D., Jones, D., Gill, S., & DeRousie, R. M. S. (2010). Implementation quality: Lessons learned in the context of the Head Start REDI trial. *Early Childhood Research Quarterly*, *25*(3), 284-298.

Douglas, K. M., & Albro, E. R. (2014). The progress and promise of the reading for

understanding research initiative. *Educational Psychology Review*, 26(3), 341–355.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research*, *18*(2), 237-256.

Engelmann, S., Hanner, S., & Johnson, G. (2002). *Corrective reading: Series guide*. McGraw-Hill College.

Fogarty, M., Oslund, E., Simmons, D., Davis, J., Simmons, L., Anderson, L., ... & Roberts, G. (2014). Examining the effectiveness of a multicomponent reading comprehension intervention in middle schools: a focus on treatment fidelity. *Educational Psychology Review, 26*(3), 425-449.

Gansle, K. A. (2005). The effectiveness of school-based anger interventions and programs: A metaanalysis. *Journal of School Psychology*, *43*, 321–341.a

Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical psychology review, 31*(1), 79-88.

Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research*, 71(2), 279-320. Chicago.

Gersten, R., Fuchs, L.S.,, Compton, D., Coyne, M., Greenwood, C., & Innocenti, M.S., (2005). Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education. *Exceptional Children, 71*(2), 149 -164.

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review*, *18,* 37-50.

Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychology Review*, 22, 254–273.

Guo, Y., Dynia, J. M., Logan, J. A., Justice, L. M., Breit-Smith, A., & Kaderavek, J. N. (2016). Fidelity of implementation for an early-literacy intervention: Dimensionality and contribution to children's intervention outcomes. *Early Childhood Research Quarterly*, *37*, 165-174.

Gwet, K. (2001). Handbook of inter-rate reliability: how to estimate the level of agreement between two or multiple raters. Gaithersburg: STATAXIS Publishing Company

Hamre, B. K., Justice, L. M., Pianta, R. C., Kilday, C., Sweeney, B., Downer, J. T., & Leach, A. (2010). Implementation fidelity of MyTeachingPartner literacy and language activities: Association with preschoolers' language and literacy growth. *Early Childhood Research Quarterly*, *25*(3), 329-347.

Harmon, J. M., Hedrick, W. B., & Fox, E. A. (2000). A content analysis of vocabulary instruction in social studies textbooks for grades 4-8. *The Elementary School Journal*, *100*(3), 253-271.

Hohmann, A. A., & Shear, M. K. (2002). Community-based intervention research: Coping with the "noise" of real life in study design. *American Journal of Psychiatry*, 159, 201–207

Horner, R. H., Carr, E.G., Halle, J., McGee, G., Odom, S., Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165-179.

International Reading Association. (2012). Adolescent literacy: A position statement of the

International Reading Association.

Justice, L. M., Kaderavek, J. N., Fan, X., Sofka, A., & Hunt, A. (2009). Accelerating preschoolers' early literacy development through classroom-based teacher–child storybook reading and explicit print referencing. *Language, Speech, and Hearing Services in Schools*, *40*(1), 67-85.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). Improving Adolescent Literacy: Effective Classroom and Intervention Practices. IES Practice Guide. NCEE 2008-4027. *National Center for Education Evaluation and Regional Assistance*.

Kazdin, A. E. (1986). Comparative outcome studies of psychotherapy: Methodological issues and strategies. *Journal of Consulting and Clinical Psychology*. 54. 95–105.

Kemple, J. J., Corrin, W., Nelson, E., & Salinger, T. (2008). The Enhanced Reading Opportunities Study: Early Impact and Implementation Findings. NCEE 2008-4015. *National Center for Education Evaluation and Regional Assistance*.

Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley.

Kintsch, W. (1974). The representation of meaning in memory. Oxford, England: Erlbaum.

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. Review of Educational Research, 86(3), 643-680

Lee, CD., & Spratley, A. (2006). Reading in the disciplines and the chal lenges of adolescent literacy. Report to the Carnegie Corporation of New York. New York: Carnegie Corporation.

Lemons, C. J., Fuchs, D., Gilbert, J. K., & Fuchs, L. S. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher*, *43*(5), 242-252.

Lemons, C. J., King, S. A., Davidson, K. A., Berryessa, T. L., Gajjar, S. A., & Sacks, L. H. (2016). An inadvertent concurrent replication: Same roadmap, different journey. *Remedial and Special Education, 37*(4), 213-222.

MacGinitie W.H., MacGinitie R.K., Maria K, Dreyer L.G., Hughes K.E. (2000) Gates-MacGinitie reading tests—4th Ed. (GMRT-4). Riverside Publishing.

McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the journal of applied behavior analysis 1991–2005. *Journal of Applied Behavior Analysis, 40*(4), 659-672.

McKeown, M. G., Beck, I. L., & Blake, R. G. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly*, 44, 218–253.

Mendive, S., Weiland, C., Yoshikawa, H., & Snow, C. (2016). Opening the black box: Intervention fidelity in a randomized trial of a preschool teacher professional development program. *Journal of Educational Psychology*, *108*(1), 130.

Miller, W. R., & Rollnick, S. (2014). The effectiveness and ineffectiveness of complex behavioral interventions: impact of treatment fidelity. *Contemporary Clinical Trials, 37*(2), 234-241.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*(3), 247-266.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*, 315–340.

Muthén, L. K., & Muthén, B. O. (2007). Mplus. Statistical analysis with latent variables. Version 8.0. Chicago.

National Association of School Psychologists. (2007). *Prevention and Intervention Research in the Schools* [Position Statement]. Bethesda, MD: Author.

National Institutes of Health. (2011). *Learning disabilities research center request for application* (FOA Number RFA-HD-12-202). Washington, DC: Author. Retrieved from http://grants.nih.gov/grants/guide/rfa-files/RFA-HD-12-202.html

National Institutes of Literacy (2007). What content-area teachers should know about adolescent literacy. Jessup, MD.

National Research Council Committee for a Review of the Evaluation Data on the Effectiveness of NSF-Supported and Commercially Generated Mathematics Curriculum Materials, Mathematical Sciences Education Board, Center for Education, Division of Behavioral and Social Sciences and Education. (2004). On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations. Washington, DC: National Academies Press.

Ness, M. K. (2016). Reading comprehension strategies in secondary content area classrooms: Teacher use of and attitudes towards reading comprehension instruction. *Reading Horizons*, *55*(1), 58.

Nunnery, J. A., Ross, S. M., & McDonald, A. (2006). A randomized experimental evaluation of the impact of Accelerated Reader/Reading Renaissance

implementation on reading achievement in grades 3 to 6. *Journal of Education for Students Placed at Risk*, 11(1), 1-18.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33-84.

Organisation for Economic Cooperation and Development (2016). Education at a glance 2016: OECD indicators. Paris: OECD Publishing.

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, *75*(6), 829.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2016). Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychological Methods*, *21*(2), 189.

Pressley, M. (2004). The need for research on secondary literacy education. In T. Jetton & J. Dole (Eds.), Adolescent research and practice (pp. 415-431). New York: Guilford Press.

Pressley, M., Symons, S., Snyder, B. L., & Cariglia-Bull, T. (1989). Strategy instruction research comes of age. *Learning Disability Quarterly*, *12*(1), 16-30.

RAND Reading Study Group. (2002). Reading for understanding: Toward an R&D program in reading comprehension. Santa Monica, CA.

Rindskopf, D. (1984). Structural equation models: Empirical identification, Heywood cases, and related problems. *Sociological Methods & Research*, *13*(1), 109-119.

Roberts, G., Vaughn, S., Beretvas, S. N., & Wong, V. (Eds.). (2016). *Treatment Fidelity in*

Studies of Educational Intervention. New York City, NY: Taylor & Francis.

Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011. *Journal of learning disabilities*, *48*(4), 369-390.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin and Company.

Shanahan, T., & Barr, R. (1995). Reading Recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *Reading Research Quarterly*, 958-996.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, *78*(1), 40-59.

Snow, C., Ippolito, J., & Schwartz, R. (2005). What we know and what we need to know about literacy coaches in middle and high schools: a research synthesis and proposed research agenda (Part 3). *Journal of Adolescent & Adult Literacy*, *49*(4), S35-S35.

Snow, C. E., Porche, M. V., Tabors, P. O., & Harris, S. R. (2007). *Is literacy enough? Pathways to academic success for adolescents.* Baltimore: Paul H Brookes Publishing.

Solis, M., Ciullo, S., Vaughn, S., Pyle, N., Hassaram, B., & Leroux, A. (2012). Reading comprehension interventions for middle school students with learning disabilities: A synthesis of 30 years of research. *Journal of Learning Disabilities*, *45*(4), 327-340.

Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2011). Intervention fidelity in special and general education research journals. *The Journal of Special Education*, 47, 3–13. doi: 0022466911419516.

Swanson, E., Wanzek, J., McCulley, L., Stillman-Spisak, S., Vaughn, S., Simmons, D., ... & Hairrell, A. (2016). Literacy and text reading in middle and high school social studies and English language arts classrooms. *Reading & Writing Quarterly*, *32*(3), 199-222.

Swanson, E., Wanzek, J., Vaughn, S., Roberts, G., & Fall, A. M. (2015). Improving reading comprehension and social studies knowledge among middle school students with disabilities. *Exceptional Children*, *81*(4), 426-442.

Swanson, E.A., Wexler, J., & Vaughn, S. (2009). Text reading and students with learning disabilities. In E.H. Hiebert (Ed.), *Reading more, reading better* (pp. 210–230). New York, NY: Guilford.

Torgeson, J. K., Houston, D. D., Rissman, L. M., Decker, S. M., Roberts, G., Vaugh, S., ... & Rivera, M. O. (2007). *Academic literacy instruction for adolescents: A guidance document from the Center on Instruction*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Toste, J. R., Fuchs, D., Fuchs, L. S., & Boon, R. T. (2013). Supporting struggling readers in high school. *Adolescent literacy*, 79-91.

National Center for Education Statistics. (2009). T*he Nation's Report Card: Reading 2015*. Washington, DC: Institute of Education Sciences, U.S. Department of Education

U.S. Department of Education, Institute of Education Sciences (2011). Request for Applications: Education Research Grants (CFDA Number: 84.305A). Washington,

DC.

Vadasy, P. F., & Sanders, E. A. (2009). Supplemental fluency intervention and determinants of reading outcomes. *Scientific Studies of Reading,* 13, 383–425.

Vadasy, P. F., Sanders, E. A., & Nelson, J. R. (2015). Effectiveness of supplemental kindergarten vocabulary instruction for English learners: A randomized study of immediate and longer-term effects of two approaches. *Journal of Research on Educational Effectiveness*, *8*(4), 490-529.

Vaughn, S., Swanson, E. A., Roberts, G., Wanzek, J., Stillman-Spisak, S. J., Solis, M., & Simmons, D. (2013). Improving reading comprehension and social studies knowledge in middle school. *Reading Research Quarterly*, *48*(1), 77-93.

Vaughn, S., Roberts, G., Swanson, E. A., Wanzek, J., Fall, A. M., & Stillman-Spisak, S. J. (2015). Improving middle-school students' knowledge and comprehension in social studies: A replication. *Educational Psychology Review*, *27*(1), 31-50.

Vaughn, S., Solís, M., Miciak, J., Taylor, W. P., & Fletcher, J. M. (2016). Effects From a Randomized Control Trial Comparing Researcher and School-Implemented Treatments With Fourth Graders With Significant Reading Difficulties. *Journal of Research on Educational Effectiveness*, *9*(sup1), 23-44.

Van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*(8), 1081-108.

Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology*, *103*(2), 455.

Wheeler, J. J., Baggett, B. A., Fox, J., & Blevins, L. (2006). Treatment Integrity A Review of Intervention Studies Conducted With Children With Autism. *Focus on Autism and Other Developmental Disabilities,* 21(1), 45-54.

Wilson, D. B., & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: evidence from meta-analysis.

Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of consulting and clinical psychology*, *49*(2), 156.